Circulating nucleic acids as biomarkers of breast cancer

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Basma Rghebi

Department of Cancer Studies

University of Leicester

September 2016

Circulating nucleic acids as biomarkers of breast cancer

Basma Rghebi

Abstract

Background: Breast cancer is the commonest cancer and the second cause of cancer death among women. Two circulating biomarkers microRNAs (miRNAs) and circulating cell-free tumour DNA (cf-DNA) are currently under active investigation as a liquid biopsy for their utility in screening and management of breast cancer.

Aims: The aim was to profile circulating miRNAs in plasma of breast cancer and compare to healthy controls to determine whether there were any miRNAs that differentially expressed in cancer. A second aim was to compare different methods to detect *PIK3CA* and *TP53* mutations in breast cancer tissues DNA and cf-DNA.

Methods: The expression level of 384 miRNAs in pooled plasma of breast cancer was compared to healthy controls using TaqMan microRNA array, followed by validation of specific miRNAs in single cancer and control plasma, using qPCR. Multiple different methods (qPCR, HRMC, ddPCR and Sanger sequencing) were compared for detection of *PIK3CA* or *TP53* mutations in breast cancer tissue and matched plasma (cf-DNA).

Results: 49 miRNAs (12.8%) showed significant expression change between pooled breast cancer plasma and controls. 11 of these were upregulated ($P<0.05$) in individual cancer plasma but there was wide variation in expression in controls. Overall, the ddPCR technology had higher sensitivity in detecting *PIK3CA* (0.1%) or *TP53* (0.01%) mutations in cell line and breast cancer tissue than that achieved with other methods and overall 18 of 26 samples had a mutation in one or other gene.

Conclusions: There was expression difference of a number of miRNAs in plasma of breast cancer compared to healthy controls, which needs further validation in other cohorts. The ddPCR was able to detect *PIK3CA* or *TP53* mutations in breast cancer tissues with higher sensitivity than the other methods tested.

Acknowledgement

First of all, I would like to thank my God for everything in my life, and I would dedicate this project to my family, especially my parents who support me all the time.

I would like to express my special thanks to Dr Jacqueline Shaw, my supervisor, for her continuing support, help, supervision and guidance throughout my research work. Also, I would further like to thank the following for their assistance in many aspects of the PhD: Dr David Guttery, Dr Karen Page, Dr David Moore, Dr Jinli Luo, Lindsay Primrose and Ricky Trigg. Finally, I would like to thanks, everyone who directly and indirectly helped me in my project.

Table of contents

List of tables

List of figures

List of abbreviations

BSMS	Breast Screening and Monitoring Study.
BRCA1	Breast cancer gene one.
BRCA2	Breast cancer gene two.
cDNA	Complementary DNA.
cf-DNA	Circulating cell-free deoxyribonucleic acid.
CNAs	Copy number aberrations.
CNVs	Copy number variations.
Ct	Cycle threshold.
ΔCt	Delta cycle threshold.
ΔΔCt	Delta delta cycle threshold.
DNA	Deoxyribonucleic acid.
DdPCR	Droplet digital polymerase chain reaction.
DCIS	Ductal carcinoma in situ.
ER	Estrogen receptor.
FFPE tissue	Formalin-fixed, Paraffin-embedded tissue.
H & E staining	Haematoxylin and Eosin staining (H&E)
HER2	Human epidermal growth factor receptor 2.
HGD	Human genomic DNA.
HRMC	High resolution melting curve.
IHC	Immunohistochemistry.
mRNA	Messenger ribonucleic acid.
MeV software	Multiple Experiment Viewer.
miR, miRNA	Micro-RNA.
NGS	Next generation sequencing.
NRC	No response to chemotherapy.

NRE	No-response to endocrine therapy.
PNA	Peptide nucleic acid.
PR	Progesterone receptor.
qPCR	Quantitative polymerase chain reaction.
RC	Response to chemotherapy.
RE	Response to endocrine therapy.
RNA	Ribonucleic acid.
RQ	Relative quantification.
SOP	Standard operating procedure.

Chapter 1. Introduction

1.1 Breast cancer

Breast cancer is the most common cancer in women and causes up to 500,000 deaths per year worldwide (Maxmen, 2012). In the UK, breast cancer accounts for 31% of all new cases of cancer among women. In 2013, one in eight women in the UK had a lifetime risk of developing breast cancer. The reported incidence has increased by 50% over the last 25 years, and around 53,700 UK women were diagnosed with breast cancer in 2013; equivalent to around 150 diagnosed each day. Even though breast cancer mortality rates continue to decrease, breast cancer remains the second most common cause of cancer-related deaths after lung cancer, with approximately 11,600 breast cancer deaths in the UK in 2012 (www.cancerresearchuk.org).

Breast cancer is a heterogeneous disease with a complex interaction of several different molecular profiles, biological behaviours, different subtypes and risk profiles. The strongest risk factor for breast cancer is age; its incidence increases with age and around 81% of breast cancers develop in women older than 50. Although breast cancer is uncommon in younger aged women, it is the most commonly diagnosed cancer in women aged under 39, where around 1,350 cases of breast cancer are diagnosed each year in women aged 35-39 in the UK (www.cancerresearchuk.org). Other important risk factors associated with developing breast cancer are those which increase oestrogen exposure (including reproductive and hormonal factors), as 75% of primary breast cancers are oestrogen receptor positive (ER+) (Nadji et al., 2005) and depend on oestrogens for their growth. There are many other risk factors for breast cancer, including early menarche, late menopause, the late birth of first child and family history of breast cancer due to genetic disorders, such as *BRCA1*/*BRCA2* mutations.

1.1.1 Development of breast cancer

Breast cancer develops as a result of an accumulation of genetic and/or epigenetic mutations, which leads to activation of oncogenes and inactivation of the tumour suppressors (Wolfson et al., 2015). Of particular importance are mutations in DNA repair genes and genes which are involved in cell cycle control, apoptosis, angiogenesis, adhesion, trans-membrane signalling and genomic stability. These genetic and epigenetic mutations lead to gene expression changes, which cause modification

and disturbance to normal growth regulatory and survival pathways in mammary epithelial cells. Hence, the cancer cells can escape from normal cellular and environmental controls, leading to tumour development and progression. In solid tumours, around 30% of the genome has an abnormal copy number, which can be as large as 10^3–10^5 per tumour (Loeb and Christians, 1996; Cheng and Loeb, 1997). One of the important causes of breast cancer are genetic (hereditary) factors (Calderon-Garciduenas et al., 2005; Ayub et al., 2014), especially in young female and its main part is the germ-line mutations in *BRCA1* and *BRCA2* genes (Kwong et al., 2011). Hereditary breast cancer follows an autosomal dominant inheritance pattern and occurs mainly as an early-onset, high intensity, and bilateral form of the disease.

Breast cancer is a heterogeneous disease of several subtypes displaying distinct differences in biological and clinical behaviour. Variations in the cellular component, growth rate and the activity of signalling pathways of the tumours are all correlated with the variation in the expression of specific subsets of genes (e.g., ER, PR, HER2 and Ki67), which provide a distinctive molecular portrait of each tumour (Perou et al., 2000).

There are different gene expression changes and gene mutations (Figure 1.1) involved in the molecular profile of breast cancer. The somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% across all breast cancers (Cancer Genome Atlas Network, 2012). The *PIK3CA* and *TP53* mutations are the most common somatic mutation in breast cancer with incidence of 36% and 37% respectively of all breast cancer (Cancer Genome Atlas Network, 2012). The somatic mutations do not occur as a single process, but they arise due to numerous mutational processes, which occurred throughout the lifetime of the cancer patient (Nik-Zainal et al., 2012; Helleday et al., 2014).

The *PIK3CA* and *TP53* mutations are the most common mutation drivers in breast cancer, as 129 mutational cancer drivers have been detected in breast cancer and the most mutated drivers were: *PIK3CA*, *TP53*, *MAP3K1*, *GATA3*, *CDH1*, etc. (Cancer Genome Atlas Network, 2012). Also, 42 mutational cancer drivers have been detected in breast cancer, and the most mutated drivers were: *TP53*, *PIK3CA*, *GATA3*, *MLL3*, *MAP3K1*, etc. (Stephens et al., 2012). Recently, all classes of somatic mutation across

exons, introns and intergenic regions (whole-genome sequences) of 560 breast cancer were analysed to understand the origins and the consequences of somatic mutations in breast cancers. They found 93-protein-coding cancer genes which probably could be driver mutations, including *TP53* gene with the highest frequency followed by the *PIK3CA* gene (Nik-Zainal et al., 2016). Another study found 12 base substitution, two insertions/deletions and six rearrangement mutational signatures present in breast cancer tissue, which has specific relationships with genes involved in transcription and DNA replication (Morganella et al., 2016).

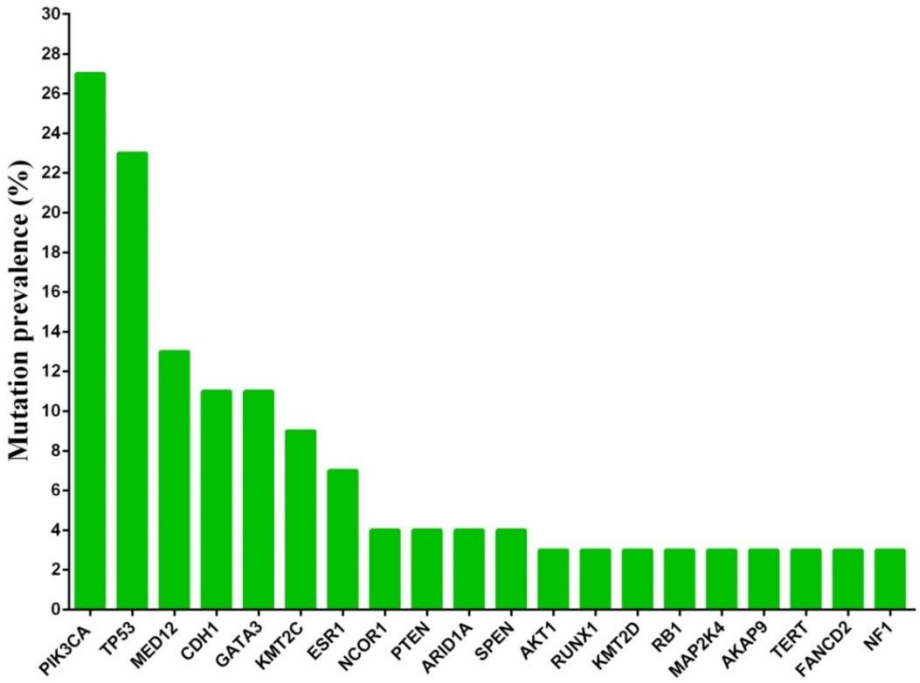


Figure 1.1. Mutated genes in breast cancer. The most frequently mutated genes in breast cancers (top 20 genes). Data obtained from the 'Catalogue of Somatic Mutations in Cancer' (www.sanger.ac.uk/COSMIC).

1.1.2 Classification of breast cancer

Breast cancer classified according to type, grade, stage and receptor status:

(1) Histopathological type: where more than 95% of breast malignancies arise from the breast epithelial tissue (breast carcinomas) of ducts or lobules and 76% of invasive breast cancers are invasive ductal carcinomas (Table 1.1).

(2) Tumour grade: classified as low (G1), intermediate (G2) and high (G3) according to the cellular morphological changes (Table 1.2).

(3) Stage of the tumour: based on the TNM classification (tumour size (T), number, size and location of lymph node involvement (N) and distant metastasis (M), (Table 1.3).

(4) Receptor status: based on the presence of oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2).

	Classification of breast cancer	% from breast cancer
Carcinoma in situ	Ductal Carcinoma in situ (DCIS)	<1%
	Lobular Carcinoma in situ (LCIS)	<1%
Invasive carcinoma	Invasive Ductal Carcinoma (IDC)	76 %
	Invasive Lobular Carcinoma (ILC)	8 %
	Lobular-Ductal Carcinoma	7 %
	Mucinous (Colloid) Carcinoma	2.4 %
	Tubular Carcinoma	1.5 %
	Medullary Carcinoma	1.2 %
	Papillary Carcinoma	1 %
	Rare Subtypes	2.9 %

Table 1.1. The Histological classification of breast cancer. The histological classification and their percentage within the breast cancer. Data obtained from (The Surveillance, Epidemiology, and End Results (SEER) Cancer Statistics Review, 1975-2011) (Howlader et al., 2014).

Grade of breast cancer		
Grade 1	Low	The cancer cells have very minimal morphological changes and grow very slowly.
Grade 2	Intermediate	The cancer cells show more morphological changes and slightly faster growth.
Grade 3	High	The cancer cells show complete abnormal morphological changes and tend to grow quickly.

Table 1.2. The breast cancer grades. The breast cancers classified as low, intermediate and high grades. Data obtained from: (www.cancerresearchuk.org).

Stages of breast cancer		
Stage 1	**Stage 1A**	The cancer is smaller than, or equal to, 2cm and has not spread outside the breast.
	Stage 1B	The cancer cells spread to the axillary lymph node and either no tumour found in the breast, **or** the tumour is 2cm or smaller.
Stage 2	**Stage 2A**	Either the tumour is < 2cm and has spread to 1 to 3 axillary lymph nodes OR it is > 2cm (but < 5cm) and has not spread to the lymph nodes OR cancer cannot be found in the breast but is in 1 to 3 axillary lymph nodes.
	Stage 2B	Either the tumour is > 2cm and < 5cm and has spread to the axillary lymph nodes OR it is > 5cm, but has not spread to the axillary lymph nodes.
Stage 3	**Stage 3A**	Either cancer cannot found in the breast, or the tumour of any size and the cancer is in 4 to 9 axillary lymph nodes OR the lump is > 5cm, and a small cluster of cancer cells has spread to the axillary lymph nodes.
	Stage 3B	Cancer has spread to tissue near the breast and may be attached to surrounding skin or muscle.
	Stage 3C	The tumour can be any size, or there may be no tumour, but it has spread to the skin of the breast, and the chest wall. It has also spread to 10 or more axillary lymph nodes or cervical or tracheal lymph nodes.
Stage 4		Cancer has spread to other parts of the body, such as the bones, liver or lungs.

Table 1.3. Stages of breast cancer. Data obtained and adapted from: (www.cancerresearchuk.org).

The main reason for breast cancer deaths is a recurrent metastatic disease. The overall 5-year survival of women with breast cancer is 99% when diagnosed at an early stage in comparison to only 15% when the disease has spread to other organs, (Table 1.4). Therefore, the early detection of breast cancer could be the key to improved overall survival.

Stages of breast cancer	Overall 5-years survival
DCIS	100% (Curative)
Stage 1	99%
Stage 2	88%
Stage 3	55%
Stage 4	15%

Table 1.4. Five-year survival of breast cancer. Five years survival statistics of Women (Aged 15-99 Years) with a different stage of breast cancer during diagnosis, between 2002-2006. Data obtained from: (www.cancerresearchuk.org).

The gene expression studies and immunohistochemical analysis have been used to classify breast cancer (molecular classification) into five subtypes depending on the presence or absence of hormone receptor expression and other specific markers (Perou et al., 2000); these subtypes are:

(i) Luminal A; ER positive and/or PR positive, low Ki67, HER-2 negative and low grade histologically (tumour grade 1 or 2).
(ii) Luminal B: ER positive can express low levels of hormone receptors, and are usually high grade. HER2-positive (or HER2-negative with high Ki67).
(iii) Basal-like breast cancer; ER negative, PR negative and HER-2 negative tumours, (e.g., Triple negative breast cancers).
(iv) HER-2 (also called ErbB2) positive tumour. ER negative and PR negative.
(v) Others; normal-breast like, molecular apocrine, and claudin low.

Treatment options (e.g., endocrine treatment, Trastuzumab and chemotherapy) depend on the five major subtypes of breast cancer. However, recently, breast cancer classifies into ten subtypes (genome-driven integrated classification) depending on joint clustering of copy number aberrations (CNAs) and gene expression profiles, (Curtis et al., 2012).

1.1.3 Screening for breast cancer

One of the most important approaches to improve prognosis is to detect cancer at an early stage (< 2 cm and node-negative), which is expected to improve the outcome and survival rate with less extensive treatment required. Studies in mouse models and human patients have shown that systemic spread of cancer cells can occur in the early pre-invasive stages of tumour progression (Husemann et al., 2008; Hanahan & Weinberg, 2011).

The Mammogram is the standard screening tool used in breast cancer screening and early detection (Nelson et al., 2009; Coresetti et al., 2011). In the UK, mammograms are used to screen all women aged between 50 and 70 and repeated every 3 years. 15,000 breast cancer cases are diagnosed annually through screening mammograms (www.cancerresearchuk.org). The specificity of screening mammography (which is the likelihood of the test being normal when cancer is absent) is 90-95%. But its sensitivity (which is the percentage of breast cancers detected in a given population when breast

cancer is present) ranges between 67% and 95%. Because it dependents on several factors, such as professional experience of the examiner, and breast density. As such, the mammogram has low sensitivity in detecting abnormal growth in high-density breast tissue (as in younger age women) (Berg et al., 2004; Checka et al., 2012). The sensitivity of screening mammogram can be as low as 40% to 50% in very dense breast tissue (Britton et al., 2012). Also, the false positive rate is high (8-10%) with mammograms (Taplin et al., 2008), and the procedure involves exposure to ionising radiation (Tabar et al., 2011). Therefore, concerns still exist for misdiagnosis, missed diagnosis and over-diagnosis (Nattinger, 2010).

In 2012, it was estimated that around 4,000 out of 15,000 women each year were diagnosed through screening mammography and have breast cancer treatment; they may have over-diagnosis and overtreatment (www.cancerresearchuk.org).

Mammograms help in detection of abnormal growth in breast tissue, but cannot differentiate between cancer and benign/harmless lumps, such as fibroma; 66-85% of suspicious mammographic lesions are non-cancerous growth (benign) (Fahy et al., 2001). Thus, diagnosis can only be confirmed with a tissue biopsy, which results in unnecessary delays for those with breast cancer and unnecessary worries for those with benign conditions, and generally increased strains on clinics and the health care system. U/S and MRI of the breast can improve cancer detection by mammography (Madjar et al., 2010; Salem et al., 2013), but they are not routinely used as screening methods for breast cancer.

With the available screening tools for breast cancer, it estimated that 10-40% of early breast cancers are missed. Furthermore, these tools are less effective in the diagnosis of breast cancers in young women, which are usually a more aggressive type of breast cancer. Moreover, these diagnostic methods cannot differentiate between malignant and benign tissues. Therefore, invasive needle or surgical biopsies are needed to confirm the presence of malignant tissue, while 60% of abnormal mammogram lesions are benign.

It is imperative to avoid overtreatment in patients, who will only experience modest benefit from the treatment while suffering from toxic side effects. On the other hand, it is also important to avoid under-treatment or incorrect treatment. Therefore, if there are biomarkers that can help with the early detection of breast cancer, this will help in early

diagnosis, treatment of the patients and optimising their individual therapy. The ideal biomarkers should be non-invasive, easily accessible, and sensitive enough to detect breast cancer early in almost all patients.

The accuracy of the previous screening tools in cancer detection may improve by combining them with blood biomarkers, such as circulating nucleic acids (miRNAs and cf-DNA), which will not be limited to high-risk women and can analysed in a relatively non-invasive manner. Blood biomarkers are expected to help in early detection of breast cancer, prognosis and predicting response to therapy.

1.1.4 Biomarkers for breast cancer

Cancer biomarkers represent a measurable characteristic which should be characterised by high specificity and sensitivity, reliability and should be easy to measure. The cancer biomarkers may be used to help either in:

(1)-Early detection of cancer (screening biomarkers): there are no screening biomarkers for breast cancer (except for radiography screening tools or manual detection). Screening biomarkers should be inexpensive, have the ability to detect primary breast cancer even in the early stage, and demonstrate high sensitivity and specificity.

(2)-Diagnosis of cancer (diagnostic biomarkers): there are no diagnostic biomarkers for breast cancer.

(3)-Treatment guidance (predictive biomarkers): There are many treatment indications for breast cancer (e.g. tumour size, stage and lymph node involvement and hormone receptor status). ER, PR and HER2 are considered as predictive biomarkers.

(4)-Prognosis of cancer (prognostic biomarkers); HER2 and Ki-67 can considered as prognostic biomarkers (Weigel and Dowsett, 2010).

The ER, PR, HER2 and Ki-67 are the only biomarkers which are approved for clinical use.

The previously known tumour markers, such as CA15.3 and carcinoembryonic antigen (CEA), are used to predict recurrence of breast cancer as they frequently elevated in

patients with late stage and metastatic disease, but they have limited specificity and sensitivity (Duffy 1999; Henry & Hayes, 2012). None of these are used for early diagnosis of breast cancer (Maric et al., 2011).

There are no long-term or specific prognostic biomarkers to identify the group of patients, who may relapse with metastatic disease in the future. Breast cancer recurrence can occur up to 20 years after the primary treatment (Karrison et al., 1999). The recurrence of tumour after a long period of remission following the primary treatment can be explained by a period of tumour dormancy (Meltzer, 1990; Murray 1995), which is common in breast cancer. The dormancy stage of cancer (micrometastasis) refers to the period where the patient is disease-free followed by recurrence of cancer. The origin of dormant cells is thought to be the circulating tumour cells (CTC) (Pantel & Speicher, 2016). These cancer cells extravasate to the secondary tumour site and then proliferate and cause tumour metastasis and progression. Alternatively, they may undergo cell death (apoptosis), or remain dormant, where they stop proliferating but survive in a quiescent state for many years (usually more than 5-6 years) before starting to proliferate again in new microenvironments, leading to recurrence of the disease.

Many studies have suggested using different blood biomarkers such as circulating nucleic acids, circulating tumour cells (Parkinson et al., 2012), and circulating tumour DNA (Dawson et al., 2013b) or autoantibodies or blood metabolomics, but none of them has been approved for clinical use.

1.1.5 Circulating nucleic acids

By definition, circulating nucleic acids are extracellular nucleic acids (DNA or RNA) circulating in the blood. Regarding their origin, they are thought to be released from cells, whether normal cells or malignant cells of primary or metastatic cancer. It is suggested that these circulating nucleic acids in the blood can be used as minimally invasive diagnostic and predictive biomarkers in benign and malignant conditions and can be used to assist in the diagnosis, monitoring and prognostication of cancer. Examples of circulating nucleic acids are cf-DNA (circulating free DNA, which consists of nuclear and mitochondrial DNA) or small non-coding RNAs, such as microRNA (miRNA).

1.1.5.1 MicroRNAs (miRNAs, miRs)

The microRNAs are a class of small, non-coding, single strand RNA molecules (~22 nucleotides long). They were discovered more than 20 years ago (Lee et al., 1993). The presence of circulating miRNAs in cell-free body fluids, such as plasma and serum of cancer patients, were reported (Lawrie et al., 2008), when they identified a high level of *miR-21* in patients with diffuse large B-cell lymphoma.

The cellular functions of miRNAs are important, as they have a negative regulatory effect on post-transcriptional gene expression, such as for tumour suppressor genes, usually by binding to their 3′untranslated region (UTR). This binding leads to blocking the translation of messenger RNA (mRNA) targets, possibly followed by mRNA degradation (He and Hannon, 2004) (Figure 1.2). This regulatory effect may vary from the weak repression of protein translation to complete cleavage of mRNA (Baek et al., 2008). Therefore, miRNAs play a role in the regulation of vital cellular processes, such as metabolism, apoptosis, cell differentiation and metastasis (Kloosterman and Plasterk, 2006). The single miRNA can bind mRNA targets; as many as 200 genes of different functions (Ambros, 2003; Zamore et al., 2005) and they are thought to regulate more than one-third of human genes. (Zamore et al., 2005).

The origin of circulating miRNAs is still not clear, whether from the tumour or other cells. Many studies suggest that blood cells are the major source of circulating miRNAs (Duttagupta et al., 2011; Pritchard et al., 2012) (Figure 1.3).

The miRNAs are present in different body fluids, such as serum, blood, saliva, urine and pleural fluid (Lodes et al., 2009; Park et al., 2009; Cortez et al., 2011). How miRNA released to the circulation is still not completely understood. But there are two hypotheses (Chin and Slack, 2008); circulating miRNA can be released into the circulation after tumour cell death and lysis (Brase, 2010), or miRNA is first released from tumour cells by active secretion (Pigati, 2010) to the tumour microenvironment, then they enter the newly-formed blood vessels (Figure 1.3). The half-life of circulating miRNA is not known, but it has been suggested that it is less than two weeks (Heneghan et al., 2011).

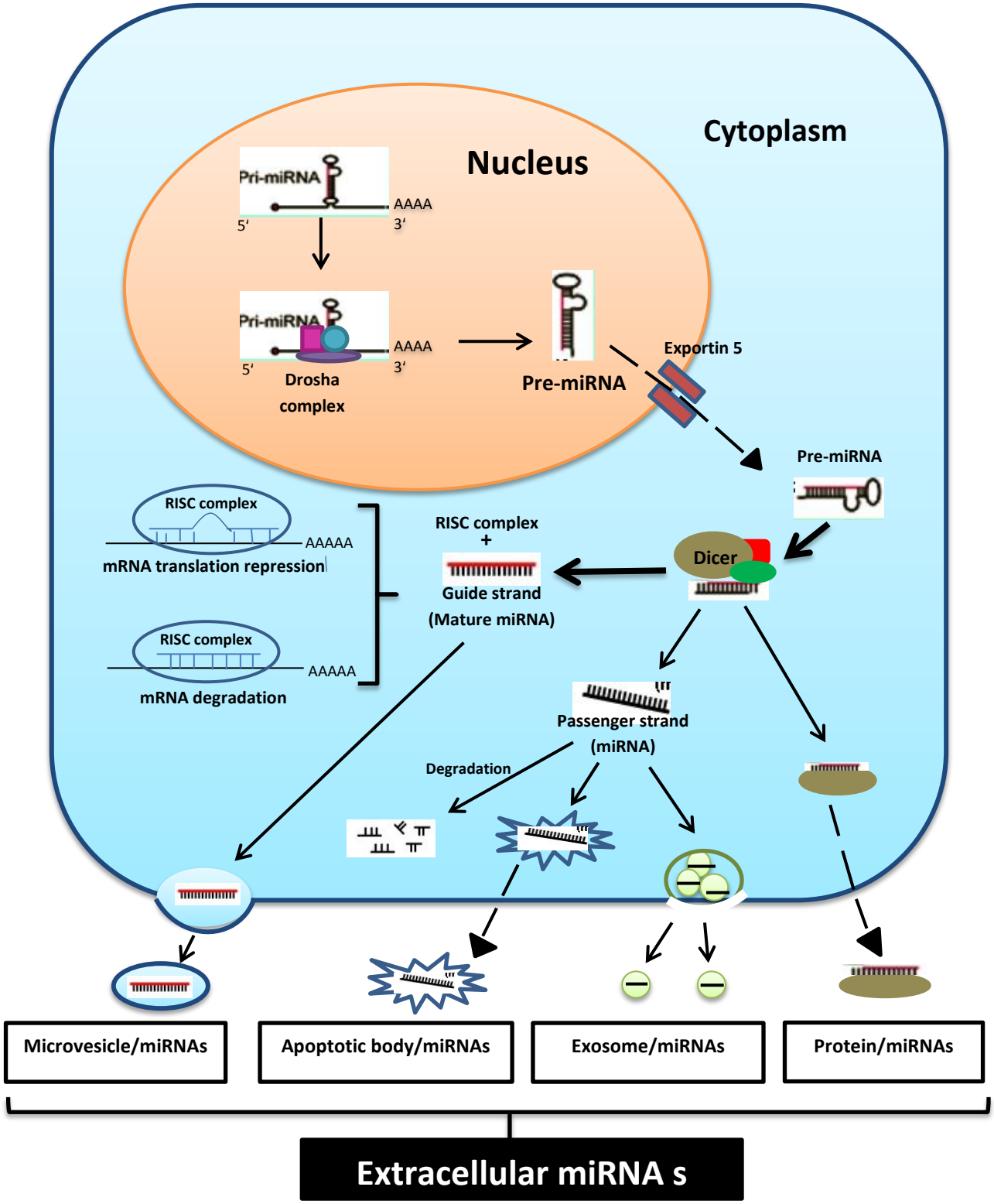


Figure 1.2. MiRNA Biogenesis, function and different route of extracellular miRNAs. The pre-miRNA passed from nucleus to cytoplasm, where it combined to RISC protein to form mature miRNA, or it digested by Dicer protein to small strand RNA (miRNA). The extracellular miRNAs excreted through inclusion in exosome, apoptotic body, microvesicle or combined with the protein. (Adapted from Zhu and Fan, 2011).

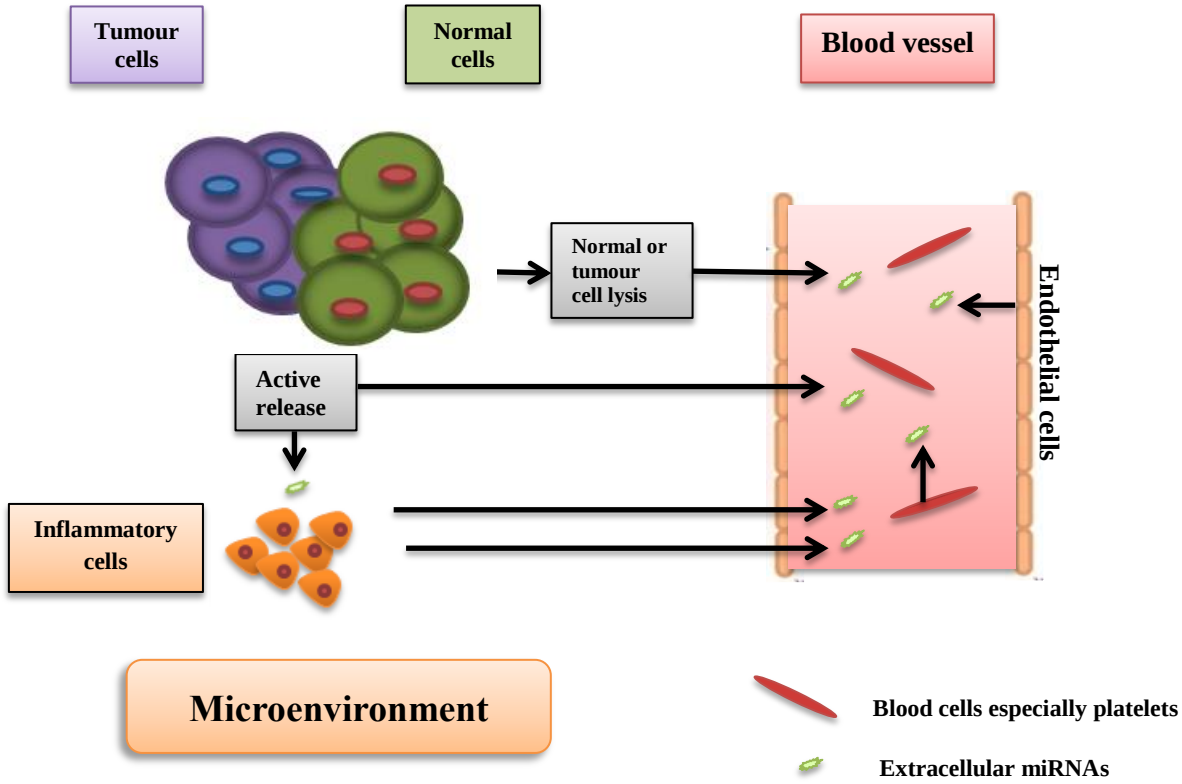


Figure 1.3. Origin of circulating miRNAs. More than 80% of circulating miRNAs originated from platelets and the rest originated from cell lysis or death of normal cells (including endothelial cells) or cancer cells or cancer microenvironment (such as stromal and inflammatory cells), which secreted first to body fluid then to the blood vessels.

The levels of miRNAs expression are different among species and tissues (Kim et al., 2004). In human tissue, more than 700 miRNAs were identified, and some of them were linked to human cancers, as they differently expressed in human malignancy compared to the matched normal tissue (Esquela-Kerscher and Slack, 2006).

Cancer initiation and progression are not completely understood, but miRNAs may play an important role. Many studies reported the involvement of miRNAs in tumourigenesis through acting as either tumour suppressors (Johnson et al. 2005) or promoters of tumourigenesis (oncogenic miRNAs) (He et al., 2005; Voorhoeve et al., 2006) or both, through post-transcriptional repression of important gene expression as tumour suppressor genes. Also, several studies reported the changes in miRNA expression in cancer tissues compared to normal tissues. These miRNAs expression differences could potentially suggest the use of miRNA as novel clinical and prognostic markers. For example, *miR-21* has been reported to be up-regulated in breast cancer tissues (Iorio et

al., 2005) and other cancers, such as glioblastoma (Chan et al., 2005) and lung cancer (Yanaihara et al., 2006) compared to normal tissues.

In human cancers, the miRNA loci are usually deleted or amplified, as they are usually presented at the fragile genomic regions, which strongly suggests the relation between miRNAs and cancer initiation and progression (Calin et al., 2004; Croce, 2009). This was supported by involvement of miRNA in almost all the human cancers, especially in breast cancer (Iorio et al., 2005; Mitchell et al., 2008; O'Day and La, 2010). Some miRNAs were up-regulated and others were down-regulated in breast cancer tissues compared to normal breast tissue. For example, 29 miRNAs differently expressed in primary breast cancer compared to normal breast tissue, where some of them were down-regulated, such as *miR-10b*, *miR-125b* and *miR-145*, while some of them were up-regulated, such as *miR-21* and *miR-155* (Iorio et al., 2005). The up-regulated and down-regulated miRNAs in the cancer patients strongly suggests that miRNA can act as tumour suppressor miRNAs and oncogenic miRNAs (Iorio et al., 2005).

Many researches have been invested in the utility of miRNAs as cancer biomarkers due to their stability in blood (Esquela-Kerscher and Slack, 2006) and their different expression in different tissues (Liang et al. 2007) and diseases. There is also the possibility to repeat measurement in a non-invasive way. The mechanism underlying the high stability of miRNA in an RNAase-rich environment is still not completely understood. However, there are two hypotheses of how miRNA are protected from degradation in circulation. The first hypothesis suggests the inclusion of miRNAs in lipid, exosomes, apoptotic bodies or lipoprotein microvesicles (Hunter et al., 2008; Mitchell et al., 2008; Kosaka et al., 2010). While, the second hypothesis suggests the inclusion of miRNA in protein complexes. Recently, the majority of circulating miRNAs were found co-fractionated with protein complexes (Arroyo et al., 2011), and bound to Argonaute proteins (Arroyo et al., 2011; Turchinovich et al., 2011), so they are stable in circulation even outside the exosome.

Many miRNAs were suggested as novel non-invasive biomarkers for early detection of different cancers, such as plasma *miR-29a* and *miR-92a* in colorectal cancer (Huang et al., 2010) and *miR-221*, *miR-376c* and *miR-744* in gastric cancer (Song et al., 2012), and plasma miR-*155* in early detection of oesophageal cancer (Liu et al., 2012).

Many studies have detected differences in miRNA expression between breast cancer patients and healthy females, either in tissue or circulation (Table 1.5). They proposed their use as diagnostic, predictive, or prognostic biomarkers. The use of miRNA profiling in whole blood was suggested as a novel method for detection of early stage breast cancer (Schrauder et al., 2012). A pilot study suggested that the changes in the levels of circulating miRNA (up- or down-regulation) may serve as biomarkers for early detection of breast cancer (Zhao et al., 2010a). Also, some miRNAs were suggested as prognostic and progression biomarkers for breast cancer, such as *miR-21* overexpression. The overexpression of *miR-21* was found to correlate with advanced tumour stage, lymph node metastasis, and poor patient survival (Yan et al., 2008). Some miRNA expression levels (e.g., *miR-9* and *miR-375*) were found to be strongly associated with ER-positive status of breast cancer (Zhou et al., 2012). Other miRNAs showed a change in their expression levels after treatment. For example, the Heneghan group (Heneghan et al., 2010 a) found a significant elevation in *miRNA-195* expression in breast cancer patients pre-operatively followed by a significant reduction after surgery in the same patient.

Higher expression of *miR-21* was observed in a different type of cancer. In breast cancer, high expression of *miR-21* observed in the serum of patients with breast cancer or benign breast diseases compared to healthy controls (Schwarzenbach et al. 2012a). Also, its higher expression was detected in primary breast cancer tissue and breast cancer cell lines (Iorio et al. 2005).

miRNA	Origin	Change in miRNA expression
miR-10b	Whole blood	No change (Heneghan et al., 2010 b).
	Whole blood	↑. In ER-ve than ER+ve, but it cannot discriminate breast cancer from healthy controls (Heneghan et al., 2010 a).
	Serum	↑. In metastatic breast cancer (stage M1) compared to normal controls (Roth et al., 2010).
	Breast cancer cell lines	Higher in micrometastatic breast cell lines BC-M1 and BC-S1 than MDA-MB231 and GI-101 (Roth et al., 2010).
	Metastatic breast cancer cells	↑. In metastatic breast cancer cells, and if it is high in non-metastatic tumour indicating invasion, metastasis and poor prognosis. It has been suggested as prognostic biomarkers (Ma et al., 2007).
miR-125b	Primary breast cancer tissue and breast cancer cell lines)	↓ (Iorio et al., 2005).

miR-145	Primary breast cancer tissue and breast cancer cell lines)	↓ (Iorio et al., 2005).
miR-21	Serum	↑ (Wang et al., 2010).
	Serum	↑ Especially stage IV cancer (Asaga et al., 2011).
	Primary breast cancer tissue and breast cancer cell lines)	↑ (Iorio et al., 2005).
	Breast cancer tissue	Higher in grade III breast cancer than in benign tumour or ER-/PR- breast cancer (Wang et al., 2010).
	Breast cancer tissue	Higher in breast cancer tissues than normal controls, especially with lymph node metastasis (Song et al., 2010).
	Breast cancer tissue	↑. In primary breast cancer tissue and advanced tumour stage, lymph node metastasis. It has been suggested as prognostic biomarkers (Yan et al., 2008).
	Serum	↑. In patients with breast cancer or benign breast diseases compared to healthy controls (Schwarzenbach et al., 2012(a)).
miR-155	Serum	↑ (Wang et al., 2010).
	Serum	↑. Can differentiate M0-patients and M1-patients from healthy controls (Roth et al., 2010).
	Serum	Higher expression in PR +ve than PR-ve tumour (Zhu et al., 2009).
	Primary breast cancer tissue and breast cancer cell lines)	↑ (Iorio et al., 2005).
miR-195	Whole blood	↑ . In all stage I to IV (Heneghan et al., 2010a).
	Whole blood	↑ then decreased in cancer patients postoperatively, to levels comparable with control subjects, following curative tumour resection (Heneghan et al., 2010a).
	Breast cancer tissue and cell lines	↓. It plays important inhibitory roles in breast cancer malignancy (Li et al., 2011).
	Breast cancer tissue	↓. In ductal carcinoma in situ (DCIS) in comparison with normal epithelium (Hannafon et al., 2011).
Let-7a	Whole blood	↑ (Heneghan et al., 2010a); Heneghan et al., 2010b),
	Breast cancer tissue	↑ then decreased in cancer patients postoperatively, to levels comparable with control subjects, following curative tumour resection (Heneghan et al., 2010a).
Let-7c	plasma	↓ In early stage breast cancer (Zhao et al., 2010a).
Let-7d	plasma	↓ In early stage breast cancer (Zhao et al., 2010a).
miR-589	plasma	↑ In early stage breast cancer (Zhao et al., 2010a).
miR-425	plasma	↑ In early stage breast cancer (Zhao et al., 2010a).
miR-34a	Serum	↑. Can differentiate M1-patients from healthy controls (Roth et al., 2010).
	Breast cancer cell line	↑ (this overexpression may be an acquired feature during carcinogenesis and cell proliferation) (Dutta et al., 2007).
miR-106a	Serum	↑ (Wang et al., 2010).
miR-141	Serum	↑ (Roth et al., 2010).

miR-126	Serum	↓ (Wang et al., 2010).
miR-199a	Serum	↓ (Wang et al., 2010).
miR-335	Serum	↓ (Wang et al., 2010).
miR-215	Serum and breast cancer cells	↑ (Van Schooneveld et al., 2012).
miR-299-5p	Serum and breast cancer cells	↑ (Van Schooneveld et al., 2012).
miR-411	Serum and breast cancer cells	↑ (Van Schooneveld et al., 2012).
miR-452	Serum and breast cancer cells	↑ (Van Schooneveld et al., 2012).
miR-202	Whole blood	↑. In early stage breast cancer (Schrauder et al., 2012).
miR-148b	Plasma	↑. In early stage breast cancer (Cuk et al., 2013).
miR-376c	Plasma	↑. In early stage breast cancer (Cuk et al., 2013).
miR-409-3p	Plasma	↑. In early stage breast cancer (Cuk et al., 2013).
miR-801	Plasma	↑. In early stage breast cancer (Cuk et al., 2013).
miR-9	Breast cancer tissue	↑. Associated with ER+ and local recurrence (Zhou et al., 2012).
miR-375	Breast cancer tissue	↑. Associated with ER+ (Zhou et al., 2012).
miR-16	Serum	↑. It has been suggested as predictive biomarkers (Hu et al., 2012).
	Plasma	↑. In breast cancer patients before chemotherapy than healthy women (Stuckrath et al., 2015).
miR-25	Serum	↑. It has been suggested as predictive biomarkers (Hu et al., 2012).
miR-222	Serum	↑. It has been suggested as predictive biomarkers (Hu et al., 2012).
miR-324-3p	Serum	↑. It suggested as predictive biomarkers (Hu et al., 2012).
	Plasma	↑. It has been significantly reduced post-surgically (Cookson et al., 2012).
miR-497	breast cancer tissue and cell lines	↓. It plays important inhibitory roles in breast cancer malignancy (Li et al., 2011).
Let-7b	Plasma	↑. Significantly reduced post-surgically (Cookson et al., 2012).
Let-7g	Plasma	↑. Significantly reduced post-surgically (Cookson et al., 2012).
miR-18b	Plasma	↑. Significantly reduced post-surgically (Cookson et al., 2012).
miR-20a	Serum	↑. In patients with breast cancer or benign breast diseases compared to healthy controls (Schwarzenbach et al., 2012(a)).
miR-214	Serum	↑. In patients with breast cancer compared to benign breast diseases and healthy controls. -Significantly reduced post- operatively. -Increased levels were associated with positive lymph node status, so it is an indicator for malignant disease and metastatic spread to regional lymph nodes (Schwarzenbach et al., 2012(a)).
miR-210	Plasma	↑. Associated with Trastuzumab sensitivity, tumour

		response and lymph node metastases (Jung et al., 2012).
miR-221	Plasma	↑. In hormone receptor negative and in patients who respond to neoadjuvant chemotherapy (Zhao et al., 2011).
miR-125b	Serum	↑. Associated with chemotherapeutic resistance (Wang et al., 2012).
miR-27a	Plasma	↑. In breast cancer patients before chemotherapy than healthy women (Stuckrath et al., 2015).
miR-132	Plasma	↑. In breast cancer patients before chemotherapy than healthy women (Wang et al., 2012).
miR-106a-5p	Plasma	↑. In pretreated patients with breast cancer compared with healthy individuals (Mishra et al., 2015).
miR-454-3p	Plasma	↑. In pretreated patients with breast cancer compared with healthy individuals (Mishra et al., 2015).
miR-495	Plasma	↓. In pretreated patients with breast cancer compared with healthy individuals (Mishra et al., 2015).
miR-195-5p	Plasma	↓. In pretreated patients with breast cancer compared with healthy individuals (Mishra et al., 2015).
miR-96-5p	Plasma	↑. In pretreated patients with breast cancer compared with healthy individuals (Matamala et al., 2015).
miR-505-5p	Plasma	↑. In pretreated patients with breast cancer compared with healthy individuals (Matamala et al., 2015).
miR-125b-5p	Plasma	↑. In pretreated patients with breast cancer compared with healthy individuals (Matamala et al., 2015).
miR-21-5p	Plasma	↑. In pretreated patients with breast cancer compared with healthy individuals (Matamala et al., 2015).

Table 1.5. The expression changes of miRNAs in breast cancer. The changes in the expression levels of different miRNAs in breast cancer tissue or circulation, which was reported previously by other research groups.

1.1.5.2 Circulating cell-free DNA (cf-DNA)

Circulating cell- free DNA (cf-DNA) is free DNA that circulates in plasma or serum. In patients with cancer, a proportion of this cf-DNA is composed of tumour-derived cf-DNA which contains tumour-specific alterations, including genetic or epigenetic changes, and is known as circulating tumour DNA (ctDNA) (Crowley et al., 2013; Diaz & Bardelli, 2014). The fractions of ctDNA within cf-DNA vary widely between 0.01 % and more than 90 % according to the tumor burden (Diehl et al., 2008). The cf-DNA in the blood detected in low concentrations in healthy individuals (Goebel et al., 2005; Umentani et al., 2006a) (approximately 10 ng/mL) and is thought to originate from apoptotic and necrotic cells. Increased cf-DNA concentrations associated with autoimmune diseases, trauma, stroke and burns (Lam et al., 2004; Chiu et al., 2006). The cf-DNA was also elevated in cancer patients (Silva et al., 1999; Ramirez et al., 2003a; Goebel et al., 2005) with levels typically > 100 ng/mL (Leon et al. 1977). The cf-DNA originates mainly from apoptosis of cancer cells, but also increased cell lysis,

necrosis and active release of cancer cells into the blood may be involved (Figure 1.4) (Stroun et al., 2001; Van Der Vaart & Pretorius, 2008). It is likely that tumour-derived and non-tumour-derived cf-DNA have distinct patterns of fragmentation, as there are differences in integrity (size of DNA fragments) between cancers and healthy controls (Wang et al., 2003). It has been hypothesised that cf-DNA which originates from tumour cell necrosis is different in length (Gormally et al., 2007), whereas cf-DNA released from the apoptotic death of non-tumour cells is uniformly digested into fragments shorter than 200 bp (Giacona et al., 1998; Jahr et al., 2001). Therefore, the integrity index, which is the ratio between the longer and shorter DNA fragments, may indicate the presence of cancer (Wang et al., 2003). The higher ratio between the long and short fragments has been reported in cancer patients (Wang et al., 2003; Jiang et al., 2006; Umetani et al., 2006a; Umetani et al., 2006b).

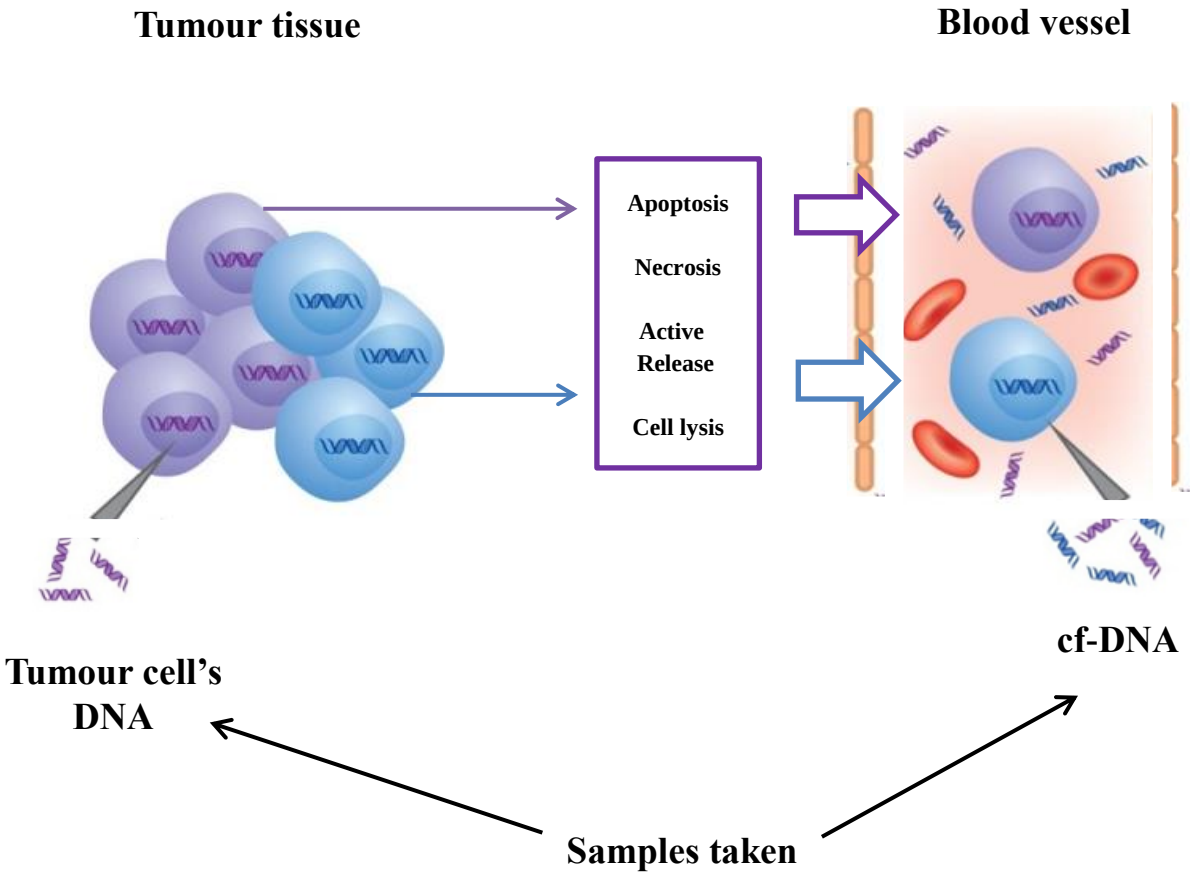


Figure 1.4. Sources of cf-DNA in cancer patients. The cf-DNA in cancer patients originates from apoptosis and/or necrosis of cancer cells, increased cell lysis and active release of cancer cells into the blood stream. (adapted from Fleischhaker and Schmidt, 2008).

Although, cf-DNA was identified over 60 years ago, its potential clinical value was not realised until the present century due to lack of sensitive and specific isolation and analytical methods. The cf-DNA can be isolated from serum, plasma and other body fluids, such as urine, sputum, breast milk, etc. (Fleischhacker and Schmidt, 2007). cf-DNA has a very short life period (1-2 hours) (Lo et al., 1999), as it is cleared rapidly from the circulation, so the cf-DNA reflects a real-time picture of the released fragments from the persistent tumour cells.

Higher levels of cf-DNA are commonly detected in malignant and benign breast cancer patients compared to the healthy controls (Zanetti-Dallenbach et al., 2007); therefore, cf-DNA concentration is not a specific marker for malignant breast cancer. Also, some studies suggest that levels of cfDNA could change after breast cancer treatment (Deligezer et al., 2008).

The concentration of cf-DNA in cancer patients is not the only difference compared to healthy individuals; promoter hypermethylation of specific cancer-related genes was found in cf-DNA in cancer patients and not in healthy individuals. However, methylation of primary tumour DNA was not often observed in cf-DNA in different cancer patients. Hence, if only one DNA methylation marker were used for diagnosis, many cancers would be missed. Also, the methylation of a certain gene promoter is not specific to a certain type of cancer, but it is a marker for cancer in general. Moreover, some DNA methylation changes can also be age-dependent in healthy individuals, so changes in DNA methylation are not necessarily indicative of tumour origin.

The other genetic and epigenetic changes of primary tumour DNA can also be detected in cf-DNA of cancer patients. For example, *Ras* oncogene (Anker et al., 1999), *TP53* mutation (Schlechte et al., 2004), *APC* mutation, microsatellite instability (MI) and loss of heterozygosity (LOH), such as loss of *cyclin D2* gene, which is suggested to be an important prognostic marker for breast cancer progression (Schwarzenbach et al., 2012b). Therefore, the same genetic characteristics of cf-DNA in cancer patients and DNA in primary cancer cells has been suggested as a diagnostic and prognostic marker of cancer (Fujimoto et al., 2004; Taback and Hoon, 2004a; Taback and Hoon, 2004b).

1.1.6 Somatic mutations in Breast cancer tissue DNA and cf-DNA

Breast cancer is a complex disease, developed as the result of the accumulation of genetic and epigenetic alterations, which leads to activation of oncogenes and inactivation of tumour suppressor in breast tissue (Wolfson et al., 2015). There are many pathways involved in breast carcinogenesis, including; (1) hormone signalling pathways (*ESR1*, *PGR* and *AR*), (2) PI3K/AKT/mTOR pathway (*AKT1*, *PIK3CA* and *PTEN*), (3) cell cycle control/DNA damage pathway (*CCND1*, *CDK4*, *CDK6*, *RB1* and *TP53*) and (4) receptor tyrosine kinase/growth factor signalling pathway (*ERBB2*/*HER2*). However, the phosphoinositide 3-kinase (PI3K) signalling pathway is the most important activated pathway in breast cancer (Samuels et al., 2004; Bader et al., 2005). Increased PI3K signals and phosphorylated AKT level were reported in many cancer such as breast, ovarian and colon cancers (Philp et al., 2001; Samuels et al., 2004).

The *PIK3CA* and *TP53* gene mutations are frequent in breast cancer (Cancer Genome Atlas Network, 2012; Stemke-Hale et al., 2008) (Figure 1.1). The data from the 'Catalogue of Somatic Mutations in Cancer' (www.cancer.sanger.ac.uk/COSMIC) reported *PIK3CA* and *TP53* mutation frequencies of 27% and 24%, respectively, in breast carcinoma tissues. These genes mutations were also reported in DCIS (Done et al., 2001b; Miron et al., 2010). The frequencies of *PIK3CA* and *TP53* mutation in DCIS were 29% and 16%, respectively (www.cancer.sanger.ac.uk/COSMIC).

1.1.6.1 Phosphatidylinositol-4,5-bisphosphate 3-kinase-catalytic sub-unit alpha *(PIK3CA)* mutation

PIK3CA mutation is one of the most common genetic alterations in breast cancers and is detected in 20–40% of these cancers (Cancer Genome Atlas Network, 2012; Stephens et al., 2012). *PIK3CA* mutation is more frequent in ER+ breast cancer (Saal et al., 2005; Liu et al., 2009) and Luminal A breast cancer subtype (Cancer Genome Atlas Network, 2012) (Table 1.6). *PIK3CA* mutations have been shown to be oncogenic in mammary epithelial cells by causing constitutive, growth factor–independent PI3K pathway

activation (Isakoff et al., 2005; Zhao et al., 2005). The *PIK3CA* gene located on the long (q) arm of chromosome 3 at 3q26.3.

Sub-types of breast cancer	Frequency of *PIK3CA* mutation (%)	Frequency of *TP53* mutation (%)
Luminal A	45%	12%
Luminal B	29%	29%
HER-2 enriched	39%	72%
Basal-like	9%	80%

Table 1.6. The frequencies (%) of *PIK3CA* and *TP53* mutations in different sub-types of breast cancer, (Cancer Genome Atlas Network, 2012).

The *PIK3CA* gene plays an important role in the Phosphoinositide 3-kinase (PI3K) signalling pathway. It has a critical role in cell growth, proliferation, survival, division and motility. The *PIK3CA* mutations are known to modulate the activity of the PI3K signalling pathway (Kang et al., 2005; Samuels et al., 2005).

PI3Ks are a family of lipid kinases which involved in signal transduction through tyrosine kinases and G-protein coupled receptors (EGF, FGF, IGF OR PDGF). These receptors are most frequently involved in cancers (Okkenhaug, 2013). *PIK3CA* is involved in the production of p110 alpha (p110α) protein, which is the catalytic subunit of the PI3K enzyme, while the other subunits produced by different genes regulate the PI3K enzyme's activity. The PI3K enzyme phosphorylates the PIP_2 protein [PtdIns (Phosphatidylinositol), PtdIns4P (Phosphatidylinositol 4-phosphate) and PtdIns (4,5)P2 (Phosphatidylinositol 4,5-bisphosphate)] to generate phosphatidylinositol 3,4,5-trisphosphate (PIP3). The PIP3 has an important role by recruiting pH domain-containing proteins to the membrane, including AKT1 and PDPK1. Also, PIP3 activates signalling pathways which involved in cell growth, proliferation, morphology, motility and survival.

PI3K signalling is a complex pathway and can be activated through interaction with tyrosine kinase receptors (Bader et al., 2005) or through the binding of its catalytic subunit to the active *RAS* (Schonleben et al., 2008). The PI3K activity can be antagonised by the action of *PTEN* (Cain et al., 2009). Because, *PTEN* catalyses the conversion of PIP_3 to PIP_2 (Engelman, 2009; Liu et al., 2009a).

The *PIK3CA* mutations in exons 9 and 20 results in amino acid changes and increase PI3K signalling pathway activity, which leads to increased cancer cell proliferation and

progression (Engelman et al., 2006). More than 80% of *PIK3CA* mutations clustered in exons 9 (helical domain) and 20 (kinase domain) with three hot spot mutations, c.1624 G>A, c.1633G>A and c.3140A>G (Cancer Genome Atlas Network, 2012; Samuels & Ericson, 2006) (Table 1.7).

Gene	Mutation (CDS)	Mutation (Amino acids)	Location	Mutation frequency	Mutation type
PIK3CA	c.3140A>G	p.H1047R	Exon-20 (Kinase domain)	55%	Nonsynonymous substitution
	c.1633G>A	p.E545K	Exon-9 (Helical domain)	20%	Nonsynonymous substitution
	c.1624G>A	p.E542K	Exon-9 (Helical domain)	11%	Nonsynonymous substitution

Table 1.7. The frequency of the commonest *PIK3CA* gene mutations in *PIK3CA* mutated breast cancer.The commonest 3 hotspots of *PIK3CA* gene mutations in breast cancer tissues and their frequency (%) in *PIK3CA* mutated breast carcinoma (www.mycancergenome.org).

The PI3K signalling pathway is one of the most frequently deregulated pathways in cancer (Yuan & Cantley, 2008). Somatic mutations in the *PIK3CA* gene, which encodes the p110*a* catalytic subunit of PI3K, were identified in human cancers in 2004 (Samuels et al., 2004) and were suggested to be mutated in approximately 15% of all tumour type (Karakas et al., 2006). *PIK3CA* mutations are the most frequent mutations in breast cancer (Kalinsky et al., 2009; Loi et al., 2010; Cizkova et al., 2012). It is also associated with development of many other cancers such as prostate cancer (Sun et al., 2009) and bladder cancer (Platt et al., 2009).

The *PIK3CA* mutations increase lipid kinase activity (PI3K pathway), which leads to increased phosphorylation of AKT and results in the transformation of normal cells to tumour cells in vivo and in vitro studies (Kang et al., 2005; Samuels et al., 2005; Gymnopoulos et al., 2007). Many studies suggest that activation of the PI3K pathway play an important role in determining breast cancer prognosis and response to chemotherapy and endocrine therapy (Engelman, 2009). However, the pathological and clinical relevance of *PIK3CA* mutations is still not completely clear. The *PIK3CA* mutations were proposed as significant and independent prognostic factors (Maruyama et al., 2007). Also, some articles suggest that the clinical outcomes and clinicopathologic findings were better with *PIK3CA* mutations; include survival benefit (Kalinsky et al., 2009; Cizkova et al., 2012). But, the prognostic significance of *PIK3CA* mutations has not been explained yet. On the other hand, *PIK3CA* driver

mutations could be one of the causes of chemotherapy resistance, but the mechanism is not clear yet (Yuan et al., 2015). Also, *PIK3CA* mutations may associate with endocrine therapy resistance (Crowder et al., 2009) or with Herceptin resistance in a cell culture model (Berns et al., 2007). However, another study showed the opposite finding (Sabine et al., 2014) and in the TEAM adjuvant endocrine study, they reported better response to treatment in patient with *PIK3CA* mutations and ER-positive breast cancer (Sabine et al., 2012). Alterations in PI3K signalling are usually involved in breast carcinogenesis as the PI3K pathway has crucial roles in cell proliferation regulation and survival (Weber, 2008). The PI3K signalling pathway can be activated by *PIK3CA* mutations or amplification. *PIK3CA* gene amplification has been reported in different cancers such as lung cancer and prostate cancer (Edwards et al., 2003; Muller et al., 2007).

1.1.6.2 *TP53* gene mutation

TP53 is a tumour suppressor gene with a large clinical impact (Vousden & Lane, 2007). It located on the short (p) arm of chromosome 17p13.1. The tumour protein p53 is a phosphoprotein composed of 393 amino acids. It consists of four domains: (1) Domain that activates transcription factors. (2) Domain that recognises specific DNA sequences (core domain). (3) Domain that is responsible for the tetramerization of the protein. (4) Domain that recognised damaged DNA, such as misaligned base pairs or single-stranded DNA.

The tumour protein p53 is a transcription factor that controls the expression of different genes. It is encoded by the tumour suppressor gene, *TP53*, which is implicated in the regulation of the cell cycle, apoptosis, DNA repair and genomic integrity. Therefore, the tumour protein p53 prevents tumorigenesis (Levine et al., 1991; Ko & Prives, 1996; Oren, 1999; Vousden, 2000; Parant & Lozano, 2003). The p53 was first discovered in 1979 (DeLeo et al., 1979; Linzer & Levine, 1979) and it exerts many functions that impact on all features of cancer (Hanahan & Weinberg, 2011). However, despite more than 35 years of research, the mechanism of how these functions interact or which function is essential to its tumour-suppressor function or the exact molecular mechanism of tumour protein p53's tumour suppressor function are still not entirely clear.

TP53 mutations distributed in all coding exons of the *TP53* gene, but the high predominance of mutations is in exons 4-9, which encode the DNA-binding domain of the protein. 30% of *TP53* mutations in exons 4-9 present within 6 "hotspot" residues (residues R175, G245, R248, R249, R273, and R282) and are frequent in almost all types of cancer (Cho et al., 1994).

The *TP53* mutation is more frequent with basal-like breast cancer (Cancer Genome Atlas Network, 2012) than the other breast cancer subtypes (Table 1.6). The most common missense *TP53* mutants in breast cancer were c.524G>A, c.818G>A and c.743G>A respectively (Table 1.8) (www.cancer.sanger.ac.uk/COSMIC). *TP53* mutations c.524G>A and c.818G>A are gain-of-function mutations (Petitjean et al., 2007).

Gene	Mutation (CDS)	Mutation (Amino acids)	Location	Carcinoma (24%)	Mutation type
TP53	c.524G>A	p.R175H	Exon-5	5.3%	Nonsynonymous substitution
	c.743G>A	p.R248Q	Exon-7	4.4%	Nonsynonymous substitution
	c.818G>A	p.R273H	Exon-8	4.1%	Nonsynonymous substitution

Table 1.8. The frequency of the commonest *TP53* gene mutations. The commonest three hotspots of *TP53* gene mutations in breast cancer tissues and their frequency percentages in breast carcinoma (www.cancer.sanger.ac.uk/COSMIC).

The *TP53* is the most commonly mutated gene in most human cancers (Hainaut & Hollstein, 2000) and it is mutated in more than 50% of all cancers; 95% of these mutations are missense mutations which alter the function and regulation of tumour protein p53 (Freed-Pastor & Prives, 2012). The most conserved area of the *TP53* gene is the core domain where the majority of mutations occur, and that may interfere with the protein–DNA interaction or affect the DNA structure stability, and lead to the synthesis of an abnormal p53 protein.

The p53 regulates DNA repair, terminal differentiation, cell growth arrest, and apoptosis in response to the oncogenic cellular stress such as DNA damage (Levine, 1997; Sigal & Rotter, 2000; Vogelstein et al., 2000). Therefore, loss of function of mutated *TP53* inhibits cell growth arrest and/or apoptosis resulting in cell proliferation without control and may lead to the initiation of carcinogenesis (Zhuo et al., 2009). Thus, the *TP53* gene plays an important role in the pathogenesis of different cancers,

including breast cancer (Greenblatt et al., 1994). Some studies reported that loss of function of p53 results in mammary carcinoma (Xu et al., 1999; Jonkers et al., 2001). Moreover, The Cancer Genome Atlas project showed *TP53* as one of the most frequent genetic alterations in breast cancer (Cancer Genome Atlas Network, 2012). *TP53* mutations have been reported in 15–71% of breast tumours. They are associated with more aggressive phenotypes, such as negative estrogen receptor status and high histological grade, compared to breast tumours without *TP53* mutations (Borresen-Dale, 2003; Olivier et al., 2006). Many studies reported the association of *TP53* mutations with a poor prognosis in breast cancer (Baker et al., 2010; Eikesdal et al., 2014). However, other studies found contradictory results, especially in patients receiving chemotherapy (Bertheau et al., 2008; Varna et al., 2011) and that could be due to the effect of *TP53* mutations on chemosensitivity of breast tumours (Bertheau et al., 2007). On the other hand, the presence of the *TP53* mutation was associated with poor prognosis in patients treated with hormonal therapy (Kim et al., 2010).

TP53 might be an early event occurring in DCIS or even earlier than DCIS stage (Ho et al., 2000; Done et al., 2001 a; Done et al., 2001 b; Kang et al., 2001; Zhou et al., 2009).

1.1.7 Copy number variations (CNVs) and aberration (CNAs)

Copy number variations (CNVs) are amplified or deleted regions of the genome, which occur in hundreds of places and cover approximately 12% of the human genome (Redon et al., 2006). These include areas coding for disease-related genes so that could be responsible for genetic changes and the susceptibility to particular diseases, such as neurological disorders, autoimmune diseases and leukaemia (Iafrate et al., 2004; Sebat et al., 2004). They range from one kilo-base to several mega-bases in size. CNVs are considered to be a major source of normal human genome variability (Iafrate et al. 2004; Sebat et al. 2004). Also, they contribute significantly to phenotypic variation (Redon et al. 2006) through their impact on gene expression and protein structure (Stranger et al., 2007; Ng et al., 2008; Stankiewicz and Lupski, 2010). About half of currently known CNVs are present in protein-coding regions (Sebat et al., 2004).

CNVs are present in healthy individuals, with an average of 12 CNVs per genome (Feuk et al., 2006). CNVs may be benign, or have an unnoticeable effect, or predispose to or cause disease in the current generation, or even in the next generation. CNVs are

thought to cause diseases through several mechanisms, depending on their proximity to structural variants so that genes could be affected in many ways. First, CNVs can directly affect the number of copies of a gene in a cell (i.e. gene dosage) by insertions or deletions of the gene, and that can cause changes in the gene expression, which may lead to genetic diseases. Deletion of the gene can also unmask a recessive mutation on the homologous chromosome, which is normally not being expressed. Second, the CNVs present within the gene sequences can decrease or inhibit the gene expression through inversions or translocations. Third, the CNVs can indirectly influence gene expression through interacting with its regulatory elements, which may lead to increased or decreased expression of the dosage-sensitive gene. Fourth, the combination of two or more CNVs may predispose to cause disease, whereas each CNV alone will cause no effect. Also, the combination of CNVs with other genetic and environmental factors might lead to complex diseases (Feuk et al., 2006).

In cancer, copy number aberrations (CNAs) are common. CNAs involving deletions have been shown as a cause of cancer susceptibility, which occur in approximately 30% of cancer-predisposing genes, such as *BRCA1*, *BRCA2*, *TP53*, *APC*, *SMAD4* and mismatch repair genes (Petrij-Bosch et al., 1997; Casilli et al., 2006). Also, germline deletions and insertions of large DNA segments have been reported as predisposing factors, which increase susceptibility to neuroblastoma (Diskin et al., 2009), colorectal cancer (Thean et al., 2010), pancreatic cancer (Lucito et al., 2007) prostate cancer (Liu et al., 2009) and *BRCA1*-associated ovarian cancer (Yoshihara et al., 2011).

Specific CNAs may be characteristic of different tumour types. The specific breast cancer-related CNAs have been described from cf-DNA and primary tumour cells of breast cancer patients (Shaw et al., 2012). Some of these CNAs were suggested to distinguish between preoperative breast cancer patients, those who have had surgery and treatment, and also between breast cancer patients during follow-up, and healthy controls (Shaw et al., 2012).

1.2 Hypothesis

The hypothesis to be tested in this thesis is that the expression of specific miRNAs in plasma and cancer-specific cf-DNA (with *PIK3CA* or *TP53* mutations) may be

significantly different between patients with breast cancer and healthy female controls. If successfully validated, these circulating miRNA and/or cf-DNA biomarkers may offer novel tools for detection and monitoring of breast cancer.

1.3 Aims and objectives

The aim was to profile circulating miRNAs in plasma of women with breast cancer and compare to healthy controls to determine whether there were any miRNAs which differentially expressed in cancer. Also, to compare different methods for detecting *PIK3CA* and *TP53* gene mutations in breast cancer tissues DNA and matched plasma (cf-DNA).

Objectives:

(1)-Identify the expression differences of 384 miRNAs between pooled plasma samples from breast cancer patients and the healthy controls, using TaqMan microRNA array cards.

(2)-Validate the cancer-specific miRNAs in individual cancer and control plasma samples, using real-time qPCR.

(3)-Validate different methods (qPCR, ddPCR and Sanger sequencing) in detecting the *PIK3CA* mutations in various concentrations of diluted cell line DNA with human genomic DNA (HGD), breast cancer tissue DNA and matched plasma (cf-DNA).

(4)-Validate different methods (HRMC, ddPCR and Sanger sequencing) in detecting the *TP53* mutations in various concentrations of diluted cell line DNA with HGD, breast cancer tissue DNA and matched plasma (cf-DNA).

(5)-Statistical analysis of all the collected data and assess the association of circulating biomarkers with radiological response to the neoadjuvant treatments.

Chapter 2. Materials and Methods

2.1 Materials

Materials	Company
Alcohol (99% IMS)	Genta Medical
Bio-Rad's PX1™ PCR plate sealer	Bio-Rad
Step One Plus ™ Real-Time PCR System Thermal Cycling Block Instrument (96 Wells)	Applied Biosystems
C1000 Touch™ Thermal cycler	Bio-Rad
Cell lines	ATCC cell lines
2x ddPCR supermix for probes	Bio-Rad
DG8™ droplet generation oil	Bio-Rad
DG8™ Cartridge	Bio-Rad
DG8™ Gasket	Bio-Rad
Digital Picture Exchange (DPX)	Fisher Scientific
Ethanol	SIGMA-ALDRICH
Eosin stain	Fisher Scientific
ExoQuick™ plus Thrombin protein kit	System Biosciences (SBI)
Foil heat seal	Bio-Rad
Forward and Reverse primers	SIGMA
GeneRead™ DNA FFPE kit	Qiagen
3730 Genetic Analyser	Applied Biosystems
Hematoxylin stain	SIGMA-ALDRICH
HGD (200ng/ µL)	Roche Diagnostics
7900HT Fast Real-Time PCR system	Applied Biosystems
Ion AmliSeqTM Library Kit 2.0	Life Technology
Ion Torrent Personal Genome Machine™ (PGM™)	Life Technology
Megaplex RT primers	Applied Biosystems
MeltDoctor™ HRMC Mastermix	Applied Biosystems
MicroAmp® Fast Optical 96-Well Reaction Plate with Barcode (0.1mL)	Applied Biosystems
MiRNA assays	Applied Biosystems
Mutant and wild-type probes, PNA and *PIK3CA* Reverse primers ligated with LNA.	Eurogentec
NaCl (1M)	SIGMA-ALDRICH
PBS solution	PAA-The cell culture company
Phenol/Chloroform/IAA	SIGMA-ALDRICH
Platinum Taq DNA high-fidelity polymerase Kit	Life Technologies
Proteinase K	Roche
QIAamp Circulating Nucleic Acid Kit	Qiagen
QIAamp® DNA mini kit	Qiagen
Qiagen miRNeasy® serum/plasma kit	Qiagen
QIAzol® lysis reagent	Qiagen
Qubit dsDNA HS Assay kit	Life Technologies
Qubit® 2.0 Fluorometer	Life Technologies

QX200 droplet generator	Bio-Rad
QX200 droplet reader	Bio-Rad
SDS buffer	SIGMA-ALDRICH
Taq Man Genotyping Mastermix	Applied Biosystems
TaqMan® Human miRNA Array cards A version 2	Applied Biosystems
TaqMan microRNA RT kit	Applied Biosystems
TaqMan Pre-Amp mastermix	Applied Biosystems
Taq Man® Universal PCR Mastermix	Applied Biosystems
Tris buffer (TE)	Fisher Scientific
USB® ExoSAP-IT® PCR Product Cleanup kit	Affymetrix
Veriti thermal cycler	Applied Biosystems
Xylene	Genta Medical

Table 2.1. Materials and instruments used in the study.

2.2 Methods

2.2.1 Patients and samples

Patient's samples were collected as a part of two different studies:

2.2.1.1 Neocent study

The Neocent study (ISRCTN: 77234840), is a neoadjuvant study comparing chemotherapy versus endocrine therapy as a neoadjuvant treatment for early stage, locally advanced ER-positive primary breast cancer. The ethical approval was given by the Leeds (East) Multi-Research Ethics Committee (MREC 07/H1306/164) and by the Korea Food and Drug Administration (Palmieri et al., 2014).

In this pilot study, 42 postmenopausal women with early stage, locally advanced ER-positive primary breast cancer were randomly assigned to either 6 cycles of FEC_{100} chemotherapy [5-flourouracil (600 mg/m²), epirubicin (100 mg/m²), cyclophosphamide (600 mg/m²)], which was given on day 1 and repeated every 3 weeks, or aromatase inhibitors hormonal therapy (letrozole, 2.5 mg daily) for a period of 18 to 21 weeks prior to surgery. The main eligibility criteria were: (1) Primary breast cancer confirmed by a histological study (biopsy) and requiring mastectomy, but the use of neoadjuvant therapy may replace mastectomy with conservative surgery. (2) ERα (Oestrogen receptor α)-positive tumour detected by immunohistochemistry (IHC). (3) Women from

menopause up to 75 years of age. (4) Consent to repeat breast tumour biopsy. (5) Written informed consent before any specific protocol procedures. Patients were seen at regular intervals for the duration of the study. Blood and plasma samples were collected before starting the treatment (day 0) (P1), 8 weeks (P2) and after completing neoadjuvant treatment (21 weeks) (P3). True cut tumour biopsies collected at baseline (day 0) (T1), 2-4 days after starting treatment (T2) and at the end of treatment (during definitive surgery) (T3). The samples were blinded for analysis. Breast imaging (breast ultrasound and breast MRI scan) were performed at baseline, day 62 and pre-operatively to assess radiological tumour response.

The baseline Neocent plasma samples (P1) divided into 4 groups depending on the type of neoadjuvant treatment [chemotherapy (C) or endocrine therapy (E)] and radiological response [response (R) or no response (NR)] and another 4 groups of their matched plasma samples after neoadjuvant treatment (Figure 2.1).

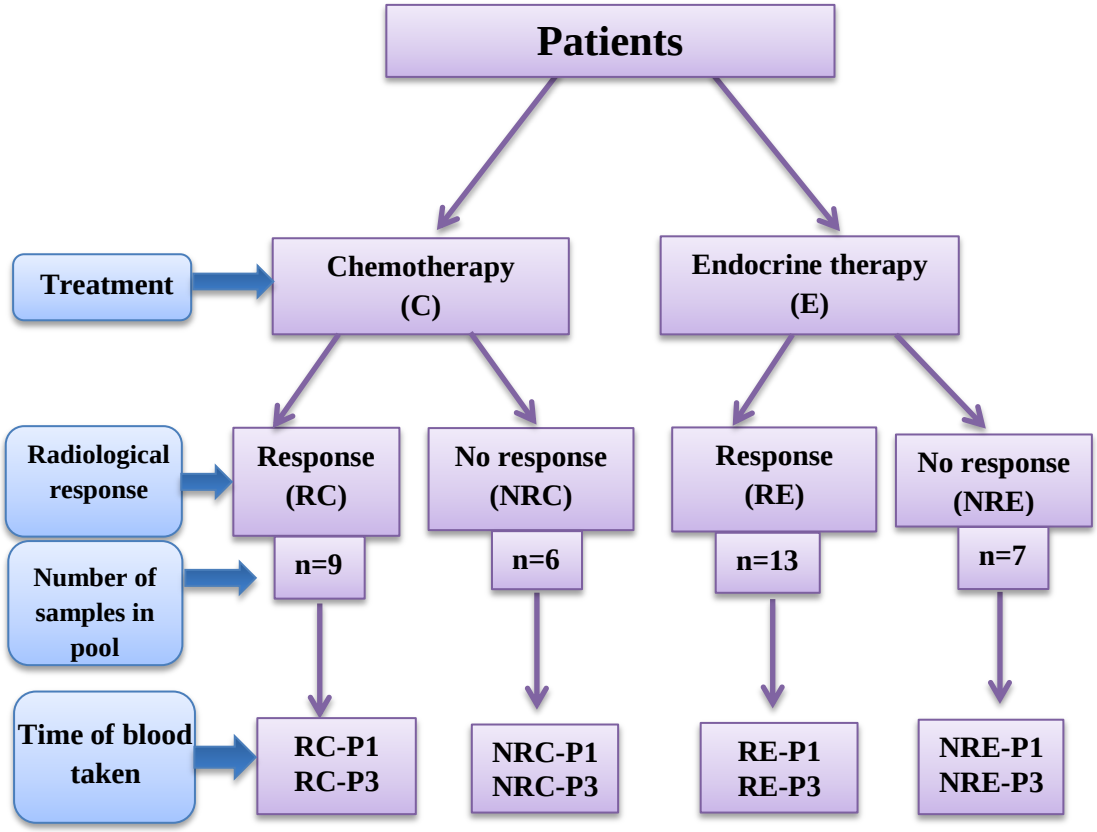


Figure 2.1. Breast cancer patient groups in the Neocent study. The diagram represents the division of Neocent patients to different groups according to treatment provided, response to treatment and time of blood sampling, before (P1) and their matched plasma samples after neoadjuvant treatment (P3). Also, a number of pooled samples in each group, which were used in a further investigation.

2.2.1.2 Breast Screening and Monitoring Study (BSMS)

The breast screening and monitoring study (BSMS) (REC: 12/LO/2019) (Appendix 3) is a study where women with an abnormal screening mammogram who were recalled for further investigations (ultrasound and fine needle biopsy) were asked to consent to give a single blood sample at the time of their follow-up appointment. Patients were recruited through a single site (Charing Cross Hospital). Approximately 1700 participants have been recruited to date, of which about 1 in 10 have biopsy confirmed invasive cancers.

2.2.1.3 Healthy Controls

17 healthy females from Leicester groups (n=5) and London group (n=12) consented to provide blood samples for the study. Also, 29 healthy females from BSMS were used as controls.

2.2.2 MicroRNA (miRNA)

2.2.2.1 The extraction of microRNAs (miRNAs) from plasma

Fresh blood was taken into EDTA-containing tubes and processed to plasma as described previously (Page et al., 2013). The blood was spun in a centrifuge at 1000 xg for 10 minutes at 4 °C to separate the plasma, which was transferred to another tube and spun at 1000 xg for 10 minutes at 4 °C to separate cell debris from plasma.

Circulating miRNA was extracted from plasma using the Qiagen miRNeasy® serum/plasma kit. QIAzol ® reagent was added to 200 µL of plasma and incubated at room temperature for 5 minutes. The samples were centrifuged for 15 minutes at 13,000 rpm at 4 °C. The upper aqueous phase containing RNA removed, and absolute ethanol added. The sample transferred to an RNeasy® Mini spin column and centrifuged at 13,000 rpm for 15 seconds, and the flow-through was discarded; the last step was repeated twice after addition of buffer RWT and buffer RPE, respectively. After that, the column was washed with RNase-free water and centrifuged at 13,000 rpm for 1 minute to elute the RNA.

2.2.2.2 The extraction of miRNA from Exosomes

The extraction of miRNAs from exosomes were to compare them to the miRNAs derived from plasma.

The ExoQuick^{TM} plus Thrombin protein kit was used to precipitate exosomes for RNA extraction, where 2.5 µL Thrombin (500 U/mL) was added to 250 µL plasma and centrifuged at 10,000 rpm for 5 minutes to precipitate the fibrin pellet.
The serum-like supernatant was transferred and treated with ExoQuick. Then it was kept at 5 °C for 30 minutes, and after that, it centrifuged at 13,000 rpm for 2 minutes to precipitate exosomes. The supernatant removed and the exosome pellet suspended in 200 µL PBS solution.
The Qiagen miRNeasy® serum/plasma kit was used to extract miRNAs from the exosome, as described in the previous section (2.2.2.1).

2.2.2.3 Reverse transcription of miRNA

The reverse transcription of miRNAs to cDNAs was previously described (Page et al., 2013). The TaqMan microRNA RT kit and Megaplex RT primers was used in reverse transcription of miRNA. 5.5 µL (100 ng/3 µL) RNA was added to 4.5 µL mastermix (consist of 0.8 µL 10x RT primers, 0.25 µL dNTPs, 1.5 µL reverse transcriptase. 0.8 µL 10x RT buffer, 1.03 µL MgCl2, and 0.125 µL RNase inhibitor) then incubated on ice for 5 minutes followed by cycling profile, using the Veriti thermal cycler (Table 2.2), to obtain cDNA. RT-negative samples (i.e. without reverse transcriptase) used as negative controls.

Temperature	Duration	Cycles
16 °C	**2 minutes**	**40 cycles**
42 °C	**1 minute**	
50 °C	**1 second**	
85 °C	**5 minutes**	

Table 2.2. PCR conditions used for reverse transcription of RNA. Temperature, duration and number of cycles used in Veriti thermal cycler for reverse transcription of RNA.

To pre-amplified cDNA, 5 µL of cDNA was added to 20 µL of mastermix (consisting of 12.5 µL TaqMan Pre-amp MM, 2.5 µL Megaplex Pre-amp primers, 5 µL nuclease free water) and incubated for 5 minutes on ice, followed by cycling profile, using the

Veriti thermal cycler (Table 2.3), then 75 µL of 0.1 Tris buffer (TE) pH 8.0 was added. RT- negative samples were used as negative controls.

Temperature	Duration	Cycles
95 °C	10 minutes	
55 °C	2 minutes	
72 °C	2 minutes	
95 °C	15 seconds	
60 °C	4 minutes	12 cycles
99 °C	10 minutes	

Table 2.3. PCR conditions used to pre-amplify cDNA. Temperature, duration and number of cycles used in Veriti thermal cycler to pre-amplify cDNA.

The pre-amplified cDNAs were checked by qPCR using TaqMan *miR-21* assays, for 40 cycles and annealing temperature = 60 °C. RT negative samples (i.e. without reverse transcriptase enzyme) and water blanks (RNase-free water) were used as negative controls, to assess genomic DNA contamination, all of which showed no amplification upon analysis by qPCR for *miR-21*. After that the cDNAs of each Neocent group were pooled and diluted to 1:100 to be analysed by TaqMan MicroRNA array.

2.2.3 DNA Extraction

2.2.3.1 DNA Extraction from Plasma

Fresh blood was taken and processed to plasma as described previously (Page et al., 2013). The QIAamp Circulating Nucleic Acid Kit used to extract cf-DNA from 1-3 mL plasma samples. Plasma samples were centrifuged at 3500 rpm for 5 minutes. Proteinase K was added to the plasma supernatant followed by 0.8 mL of ACL buffer, then incubated at 60 °C in a water bath for 30 minutes. 1.8 mL of lysate-Buffer ACB was added and then incubated on ice for 5 minutes. Transfer the samples to QIAamp Mini columns and 600 µL Buffer ACW1 was added and centrifuged followed by discard of the solution. The previous step was repeated after adding 750 µL Buffer ACW2 and then 750 µL of ethanol. The columns were centrifuged at 14000 rpm for 3 minutes and then transferred to the new collection tube and incubated at 56 °C for 10 minutes. 50 µL buffer AVE was added and incubated at room temperature for 3 minutes, and then centrifuged at 14000 rpm for 1 minute to elute nucleic acids.

2.2.3.2 DNA Extraction from Cell lines

DNA samples used for assay development and validation were from different cell lines (Table 2.4), which obtained from ATCC, except DNA from RERF-LC-Sq1 (squamous cell lung cancer) cell line was provided by PAA-The cell culture company.

DNA was extracted from cell lines pellets, which previously prepared in our lab by Lindsay Primrose. DNAs of positive cell lines for *PIK3CA* and *TP53* mutations (Table 2.4) were extracted using the QIAamp® DNA mini kit, as previously described (Page et al., 2006; Shaw et al., 2011).

Cell lines	Gene name	CDS mutation	AA mutation	Exon	Mutation zygosity	Type
HCT-116 (Primary Carcinoma of colon)	*PIK3CA*	c.3140A>G	p.H1047R	20-exon	Heterozygous	Non-synonymous substitution-Missense
MCF7 (Primary breast carcinoma, ER hormone+)	*PIK3CA*	c.1633G>A	p.E545K	9-exon	Heterozygous	Non-synonymous substitution-Missense
SW948 (Primary adenocarcinoma of colon)	*PIK3CA*	c.1624G>A	p.E542K	9-exon	Homozygous	Non-synonymous substitution-Missense
CCRF-CEM (Acute lymphoblastic leukaemia)	*TP53*	c.524G>A	p.R175H	5-exon	Heterozygous	Non-synonymous Substitution-Missense
	TP53	c.743G>A	p.R248Q	7-exon	Heterozygous	Non-synonymous Substitution-Missense
HT29 (Large intestine carcinoma)	*TP53*	c.818G>A	p.R273H	8-exon	Homozygous	Non-synonymous Substitution-Missense

Table 2.4. Cell lines used for mutation analysis. Cell lines with different *PIK3CA* or *TP53* mutations, their location, mutation zygosity status and mutation type.

Cell pellets were re-suspended in 200 µL PBS. Protease and buffer AL were added and incubated at 56 °C for 20 minutes then absolute ethanol was added. The sample was transferred to a spin column and centrifuged at 8,000 rpm (6000 xg) for 1 minute and the flow through was discarded; then buffers AW1 was added and spin at 8,000 rpm (6000 xg) for 1 minute, followed by buffer AW2 and spin at 14000 rpm (20,000 xg) for

3 minutes. After that, the buffer AE (200 µL) was added and the column stored at room temperature for 5 minutes and centrifuged at 8,000 rpm (6000 xg) for 1 minute to get the eluted DNA.

2.2.3.3 DNA extraction from breast cancer tissues

2.2.3.3.1 Haematoxylin and Eosin stain (H&E)

The formalin-fixed, paraffin embedded (FFPE) Neocent trial sections were stained with haematoxylin and eosin stain (H&E) to identify the cancer tissue from normal breast tissue. The formalin-fixed, paraffin-embedded (FFPE) Neocent trial sections (n=35) were incubated for 10 minutes at 65 °C, dewaxed in xylene and rehydrated in decreasing concentrations of alcohol (IMS 99% then IMS 95%). After that, the slides were counterstained in haematoxylin and eosin stain (H&E) and washed in running water, then dehydrated by immersion in increasing concentrations of alcohol (IMS 95% then IMS 99%), placed in xylene and mounted in DPX (Digital Picture Exchange). All H&E stained slides were reviewed by a pathologist to mark the invasive cancer area, and the foci of tumour cells were isolated by manual microdissection.

Two different methods of extraction [Phenol/Chloroform/IAA extraction and the GeneRead™ DNA FFPE kit (Qiagen Kit)] were validated to extract good quality DNA which was then used for Next generation sequencing (NGS) (Figure 2.2).

Seven FFPE slides containing the same metastatic breast cancer tissue were used to compare extraction of the tissue DNAs from FFPE tissue by two different methods, Phenol/Chloroform/IAA extraction and Qiagen kit. Three slides were used for each method, and one slide was stained with H&E and reviewed by a pathologist to mark the tumour tissue area.

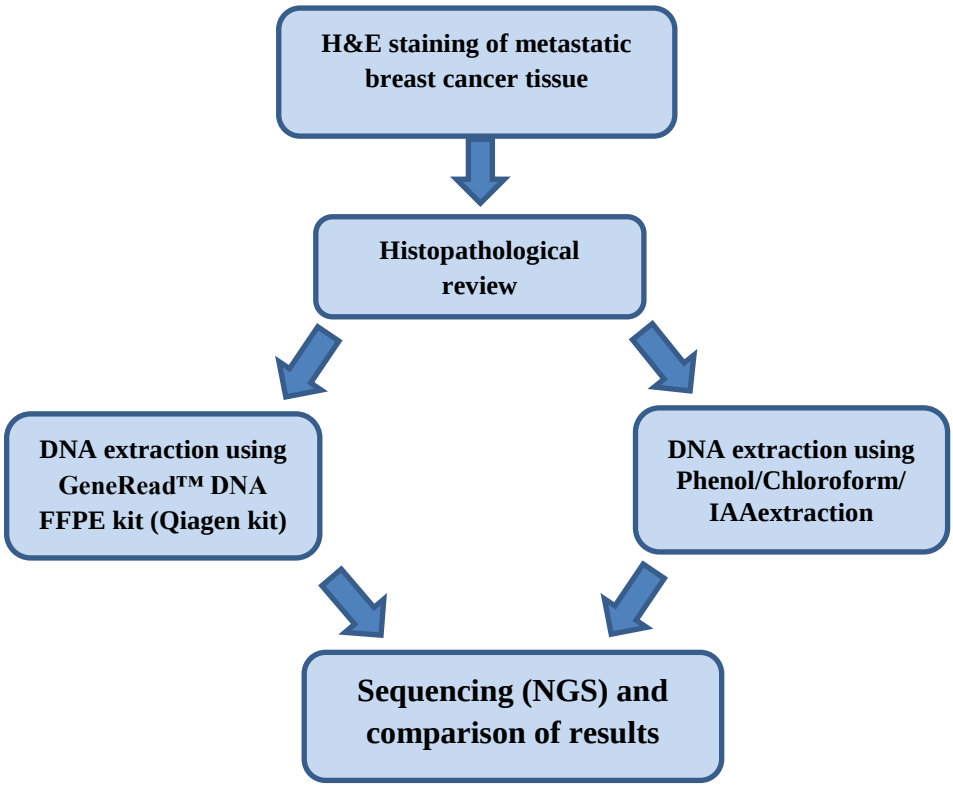


Figure 2.2. Workflow of DNA extraction steps by different methods and DNA sequencing.

2.2.3.3.2 Phenol/Chloroform/IAA method of DNA extraction

Three FFPE slides were dewaxed in xylene and rehydrated in decreasing concentrations of alcohol (IMS 99% then IMS 95%). The 0.05 M Tris pH 8.0/0.1% SDS buffer (500 µL) was added to slides and the tissue was scraped off. Tissue from each slide was pooled into one tube. Proteinase K (10 mg/mL) (25 µL) was added and tissue incubated at 55 °C for 3 nights where again 25 µL proteinase K were added every day. An equal volume (575 µL) of P:C:IAA (25:24:1) added and spin for 2 minutes. The top layer was removed, and then 1/10th total volume of 1M NaCl and 3 x volume of cold ethanol was added and incubated at -20 °C overnight. The solution was spin for 10 minutes, the supernatant layer discarded, and washed with 70% ethanol, then re-spin for 5 minutes. The ethanol was discarded, the pellet air-dried, and the pellet was then suspended in 30 µL sterile water. DNA quantity was measured (after 1:5 dilutions) using the *GAPDH*-100 bp standard curve.

2.2.3.3.3 GeneRead™ DNA FFPE kit (Qiagen kit)

The other three FFPE slides were dewaxed in xylene and rehydrated in decreasing concentration of alcohol (IMS 99% then IMS 95%). The buffer FTB (25 µL), 20 µL

proteinase K and 55 µL RNase-free water were added to the slides and tissue was scraped off. The tissue was pooled from the slides in one tube. The solution was incubated for 1 hour at 56 °C followed by another hour at 90 °C. The lower clear phase was transferred to the new tube where Uracil-N-glycosylase (UNG) and RNase-free water were added and incubated for a further hour at 50 °C; then RNase A was added. Buffer AL (250 µL) and 250 µL ethanol were added and mixed. The entire lysate was transferred to a QIAamp MinElute column and centrifuge at 14,000 rpm (20,000 xg) for 1 minute, the collection tube was discard, then 500 µL buffer AW1, followed by 500 µL buffer AW2 then 250 µL absolute ethanol (96-100%) were added, with centrifugation at 14,000 rpm (20,000 xg) for 1 minute in between. The column was centrifuged at 14,000 rpm (20,000 xg) for 1 minute, then 30 µL buffer ATE was added, incubated at room temperature for 5 minutes, then centrifuged at 14,000 rpm (20,000 xg) for 1 minute. DNA quantity was measured (after 1:5 dilutions) using the *GAPDH*-100 bp standard curve.

2.2.4 DNA quantification

One of the two following methods were used to quantified the extracted DNAs**:**

2.2.4.1 DNA quantification using a Standard Curve

The *GAPDH* in the form of the *GAPDH*-100 bp assay (Table 2.5) was used as the housekeeping gene for the quantification of the cell line and breast cancer tissue DNAs, as previously described (Shaw et al., 2011). A standard curve started with human genomic DNA (HGD 200 ng/µL) (20 ng/3.6 µL) 1:2 dilution for 7 series. It used to quantify the extracted cell line DNA through qPCR for 40 cycles and annealing temperature = 60 °C, using the Applied Biosystems Step One Plus™ Real-Time PCR System Thermal Cycling Block machine version 2.1 (Figure 2.3).

Gene	Sequence
***GAPDH*-100-FP**	5' GGCTAGCTGGCCCGATTT 3'
***GAPDH*-100-RP**	5' GGACACAAGAGGACCTCCATAAA 3'
***GAPDH*-Probe (FAM/MGB)**	5'ATGCTTTTCCTAGATTATTC 3'

Table 2.5. The *GAPDH* assay sequences. FP; Forward primer, RP; Reverse primer.

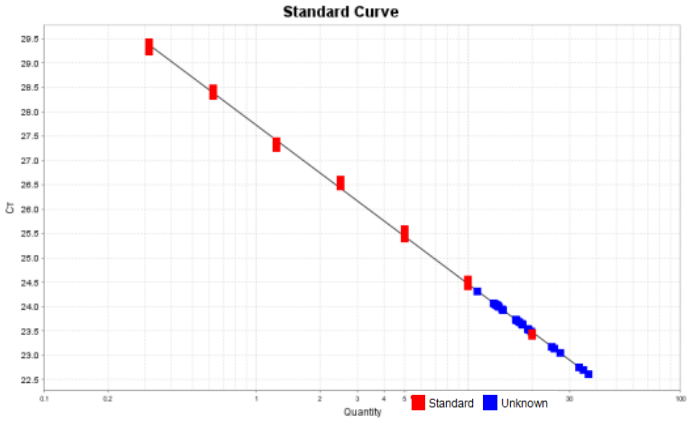


Figure 2.3. A standard curve of *GAPDH* gene. Quantification of cell line DNAs (blue dots) using the standard curve of *GAPDH*-100 (red dots). $R^2 = 0.998$, Eff%= 102.5.

2.2.4.2 DNA quantification using Qubit Fluorometer

The extracted DNAs and library prepped DNAs for NGS (Next Generation Sequencing) were quantified using Qubit® 2.0 Fluorometer with Qubit dsDNA HS Assay kit. The Qubit working solution was prepared with 199 µL assay dilution buffer and 1 µL dye reagent per Qubit assay tube. There were two tubes for standards (standard 1 and standard 2). 10 µL of each standard was added to 190 µL of the working solution and 1-20 µL of DNA sample was added to working solution to make total volume =200 µL, and then vortex tube and incubate at room temperature for 2 minutes. After that, the DNA quantification was measured using Qubit® 2.0 Fluorometer, after measuring of the standards.

2.2.5 DNA Sequencing using Ion Torrent PGM (Next generation sequencing)

Two different methods of DNA extraction of breast cancer FFPE tissue [Phenol/Chloroform/IAA extraction and the GeneRead™ DNA FFPE kit (Qiagen Kit)] were validated to extract good quality DNA which was then used for DNA sequencing using Next generation sequencing (NGS). DNA was sequenced to detect any nucleotide base difference in DNA structure (mutations, deletion or gain).

The following steps were carried out to generate DNA library with Ion Torrent Cancer Panels for sequencing with Ion Torrent PGM:

2.2.5.1 Multiplex amplification of DNA with Ion AmpliSeq™ Cancer Hotspot Panel v2

DNAs (10 ng/12 µL) which were extracted by different previous methods were added to 4 µL of 5x Ion AmpliSeq™ HiFi Mastermix and 4 µL 5x Ion AmpliSeq™ CHPv2 panel, then the PCR thermal cycler was used to amplify the target genomic regions (Table 2.6).

Step	Temperature	Time	Cycle
Enzyme activation	99 °C	2 minutes	
Denaturation	99 °C	15 seconds	17 or 22*
Annealing and Extension	60 °C	4 minutes	

*17 cycles for genomic or lymphocyte DNA and 22 cycles for cancer tissue or plasma DNA.

Table 2.6. PCR thermal cycler steps for multiplex amplification of DNA. PCR thermal cycler steps for multiplex amplification of DNA with Ion AmpliSeq™ Cancer Hotspot Panel v2.

2.2.5.2 Partial digestion of PCR products

The FuPa reagent (2 µL) was added to each amplified sample, and then the PCR thermal cycler was used for partial digestion of PCR products (Table 2.7).

Step	Temperature	Time
Enzyme activation	50 °C	10 minutes
Denaturation	55 °C	10 minutes
Annealing and Extension	60 °C	20 minutes
Final incubation	10 °C	Hold (for up to 1 hour)

Table 2.7. PCR thermal cycler steps for partial digestion of PCR products.

2.2.5.3 Ligation of adaptor

The switch solution (4 µL) and diluted barcode adaptor (2 µL) were added to each digested samples, then 2 µL of DNA ligase was added to each sample, and the thermal cycler was used to ligate the adaptor (Table 2.8).

Enzyme activation	Temperature	Time
Denaturation	22 °C	30 minutes
Annealing and Extension	72 °C	10 minutes
Final incubation	10 °C	Hold (for up to 1 hour)

Table 2.8. PCR thermal cycler steps for adaptor ligation.

2.2.5.4 Purification of DNA libraries with AMPure beads

The AMPure®XP beads reagent (45 µL) was added to each ligated sample and incubated at room temperature for 5 minutes, the tube was placed in a magnetic rack until solution clears; then the supernatant was discarded. The pellet was washed with 70% ethanol twice.

2.2.5.5 Quantification and preparation of the 100 pM libraries

The Platinum®PCR SuperMix High Fidelity (50 µL) and 2 µL of Library Amplification Primer Mix were added and mixed with each pellet and the tube returned to the magnet. Then, 50 µL of supernatant was transferred to the thermal cycler (Table 2.9).

Step	Temperature	Time	Stage
Enzyme activation	98 °C	2 minutes	
Denaturation	98 °C	15 seconds	7 cycles
Annealing and Extension	60 °C	1 minute	
Final incubation	10 °C	∞	Hold

Table 2.9. PCR thermal cycler steps for to amplify DNA libraries.

2.2.5.6 Library purification

The AMPure®XP beads reagent (25 µL) was added to each ligated sample, incubated at room temperature for 5 minutes, then the tube was placed in a magnetic rack until the solution clears. The supernatant was transferred to a new tube where 60 µL AMPure®XP beads was added and mixed with supernatant. The tube was placed in a magnet rack until the solution cleared, and then the supernatant was discarded. The pellet was washed with 70% ethanol twice. The low TE (50 µL) was added and mixed with each pellet. The tube was returned to the magnet, and the supernatant transferred to a new tube. The Qubit Fluorometer was used to measure the concentration of the stock library. For template preparations, each library prepared sample (3 µL) diluted to 0.013 ng/µL.

2.2.5.7 Sequencing with Ion Torrent PGM

Each library prepared sample (10 µL) was pooled to form combined libraries (6 DNA libraries per pool), then diluted to 100 pM libraries and followed by sequencing of DNA-template Ion Sphere Particles (ISPS) using Ion Torrent Personal Genome

Machine™ (PGM™). The DNA sequencing data were analysed using the Torrent Browser as described previously (Page et al. 2013).

2.2.6 MicroRNAs expression

2.2.6.1 TaqMan MicroRNA Array

Groups of cDNA samples were pooled together according to the type of treatment received (chemotherapy and endocrine therapy) and the response to the treatment, based on radiological response (Figure 2.1; Table 2.10). The pooled cancer samples were compared with pooled plasma miRNA samples from healthy female controls.

Cancer samples		
Time of sample	**Name of group**	**Number of pooled samples**
Baseline (P1)	Response to chemotherapy (RC)P1	n=9
	No response to chemotherapy (NRC)P1	n=6
	Response to endocrine therapy (RE)P1	n=13
	No response to endocrine therapy (NRE)P1	n=7
After neoadjuvant treatment (P3)	Response to chemotherapy (RC)P3	n=9
	No response to chemotherapy (NRC)P3	n=6
	Response to endocrine therapy (RE)P3	n=13
	No response to endocrine therapy (NRE)P3	n=7

Table 2.10. Pooled samples of breast cancer patient from Neocent study. The baseline cancer plasma samples (P1) were divided into 4 groups according to type of treatment received (chemotherapy and endocrine therapy) and their response to the treatment (based on radiological results) and another 4 groups of their matched plasma samples after treatment (P3).

TaqMan microRNA array cards were used to analyse miRNA profiles in the previous groups of Neocent pooled plasma samples. Also, a single plasma sample from each group was examined in separate TaqMan microRNA array cards to investigate the correlation between pooled and individual results. 9 µL of diluted (1:100) pre-amplified cDNAs pooling samples was added to 450 µL TaqMan Universal PCR mastermix and 441 µL nuclease free water, then 100 µL of sample was loaded in 8 holes in miRNA cards A version 2, which contain the primers of different miRNAs. The card was centrifuged for 2 minutes at 1200 rpm and run on the 7900HT Fast Real-Time PCR system. The results were analysed using SDS software version 2.2.2.

The cDNAs pre-amplified from a pooled sample of healthy females (Leicester groups n=5 and London group n=12) were used as a control.

TaqMan MicroRNA Arrays for cDNAs of individual samples {N22 (CR), N3 (NCR), N8 (ER) and N7 (NER)} from each group were performed.
Delta cycle threshold (ΔCt), delta delta Ct (ΔΔCt) and relative quantitation (RQ) were calculated for each miRNA, (Table 2.11).

ΔCt	$Ct_{\text{of each miRNA}}$ - average $Ct_{\text{of 384 miRNAs (exclude miRs with Ct>35)}}$
ΔΔCt	$\Delta Ct_{\text{of each miRNA in cancer sample}}$ - $\Delta Ct_{\text{of the same miRNA in control sample}}$
Fold change	**-ΔΔCt**
RQ	$2^{-\Delta\Delta Ct}$

Table 2.11. The calculation of ΔCt, ΔΔCt and RQ. The calculation of ΔCt, ΔΔCt and RQ for TaqMan microRNA data analysis.

2.2.6.2 TaqMan qPCR of selected miRNA

The validation of cancer-specific miRNA expression changes, which were detected by the MicroRNA Array in pooled samples, was through qPCR assays on pre-amplified cDNA of each Neocent sample (35 different cancer samples) and 17 control plasma samples [Leicester controls (n=5), London controls (n=12)].

The diluted pre-amplified cDNA (1:100) (4.5 µL) was added to 5.5 µL mastermix (consist of 5 µL TaqMan® Universal PCR mastermix, 1.5 µL RNase-free water and 0.5 µL specific miRNA target primers) then loaded in triplicate for each sample in a 96-well plate. A water blank (RNase-free water) and RT-negative samples were used as negative controls [no template control (NTC)]. While, both miRNAs (*miR-16* and *miR-484*) were selected to be used as endogenous positive controls (reference genes) due to their stable expression profile in almost all plasma samples in our study. They were the most stable miRNAs by 2 ways; manually through calculation of standard deviation (SD) of delta Ct of the same miRNA in control and cancer samples, then chose miRNAs with smallest SD and they were *miR-16* and *miR-484*. Another way was to use the GeNorm programme (Vandesompele et al., 2002), where the RQ is calculated for controls and cancer samples (Table 2.11). After that, RQ values were loaded in GeNorm programme which chose *miR-16* and *miR-484* as the most stable miRNAs. The *miR-16* expression was stable and reproducible across all the blood samples in other sudy and therefore used to normalize RQ-PCR data (Henegan et al., 2010 b). Also, other study suggest the use of the combination of miR-484/miR-191 as endogenous controls (Hu et al., 2012). However, other study found miR-16 normalisation was not justifiable (Cookson et al., 2012).

The Applied Biosystems StepOne Plus™ Real-Time PCR System Thermal Cycling Block machine with TaqMan® reagent was used to detect the expression of the specific miRNA target for 50 cycles (Table 2.12), as described previously (Page et al., 2011). The FAM was used as a reporter and NFQ-MGB as quencher.

StepOne software version 2.1 with 0.1 as detected threshold was used to identify the cycle of miRNA expression which was called the cycle threshold (Ct); it also measured Ct Mean and standard deviation (SD) of triplicate samples.

Delta Ct (ΔCt) of each selected miRNA in cancer samples was calculated using the average Ct of *miR-16* and *miR-484* as reference genes.

Steps	Temperature	Time	Cycles
Initial denaturation	**95 °C**	**20 seconds**	
Denaturation	**95 °C**	**1 second**	**50**
Annealing/Extension	**60 °C**	**20 seconds**	

Table 2.12. The qPCR conditions used for cDNA. Steps, temperatures and time periods for each cycle in the qPCR for cDNA.

2.2.6.3 Statistical analysis of microRNA array data

The significant difference in the expression of 384 miRNAs in pooled plasma samples of cancer patients and pooled controls were statistically analysed using GraphPad Prism version 6 (multiple t-test-one per row) with a significance of *P*-value <0.05 and MeV (Multiple Experiment Viewer) software method to identify the cancer-specific miRNAs (i.e. significant up- or down-regulated miRNAs in cancer patients compared to female healthy controls). Statistical analysis of the correlation between pooled and individual samples results was performed used Spearman test (r) with a significance of $P<0.05$. The significant difference in the expression of cancer-specific miRNAs in the individual plasma or exosome sample of cancer patients and controls were statistically analysed using GraphPad Prism version 6, un-paired t-test (Mann-Whitney test) or one-way ANOVA (Tukey test) with a significance *P*-value <0.05.

2.2.7 Somatic Mutations in Breast cancer tissue DNA and cf-DNA (*PIK3CA* and *TP53* mutations)

The top 3 hotspots of *PIK3CA* mutations in the breast cancer tissues were *PIK3CA* c.3140A>G, c.1633G>A and c.1624G>A respectively (Cancer Genome Atlas Network,

2012). The top 3 hotspots of *TP53* mutations in breast cancer tissues were c.524G>A, c.818G>A and c.743G>A respectively (www.sanger.ac.uk/COSMIC). The presence of these common hotspots of *PIK3CA* and *TP53* mutations were chosen to examine in the cancer tissue DNA and the corresponding cf-DNA to find if they can potentially be used as biomarkers to detect the breast cancer.

The mutation assays (Forward, reverse primers and specific allele probes) were designed using Primer 3 and Primer Express programmes. In cancer tissue samples, somatic mutations usually present in very small amounts compared to a large amount of wild-type DNA. Detection of these low mutation fraction in the presence of a high background of wild-type alleles can be challenging. Therefore, the wild-type allele was blocked using a Peptide Nucleic Acid (PNA) (Figure 2.4), which was designed by the PNA BIO online programme.

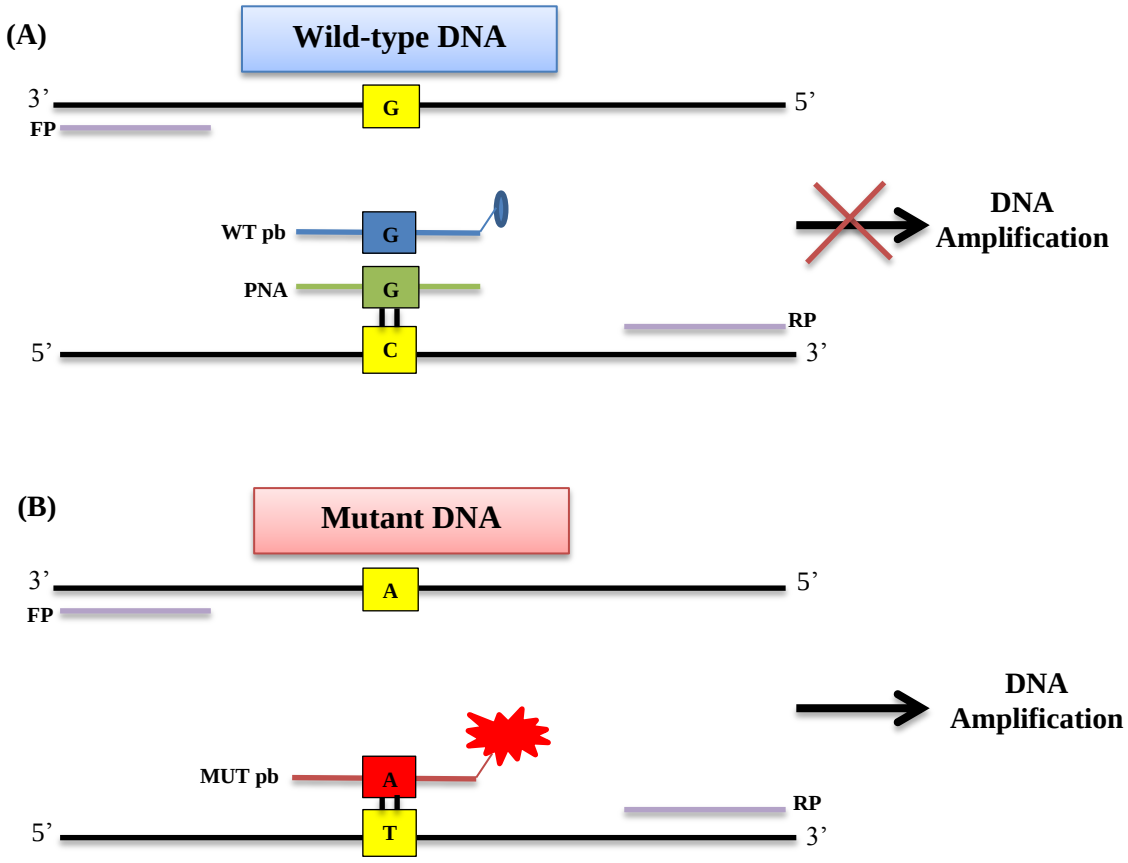
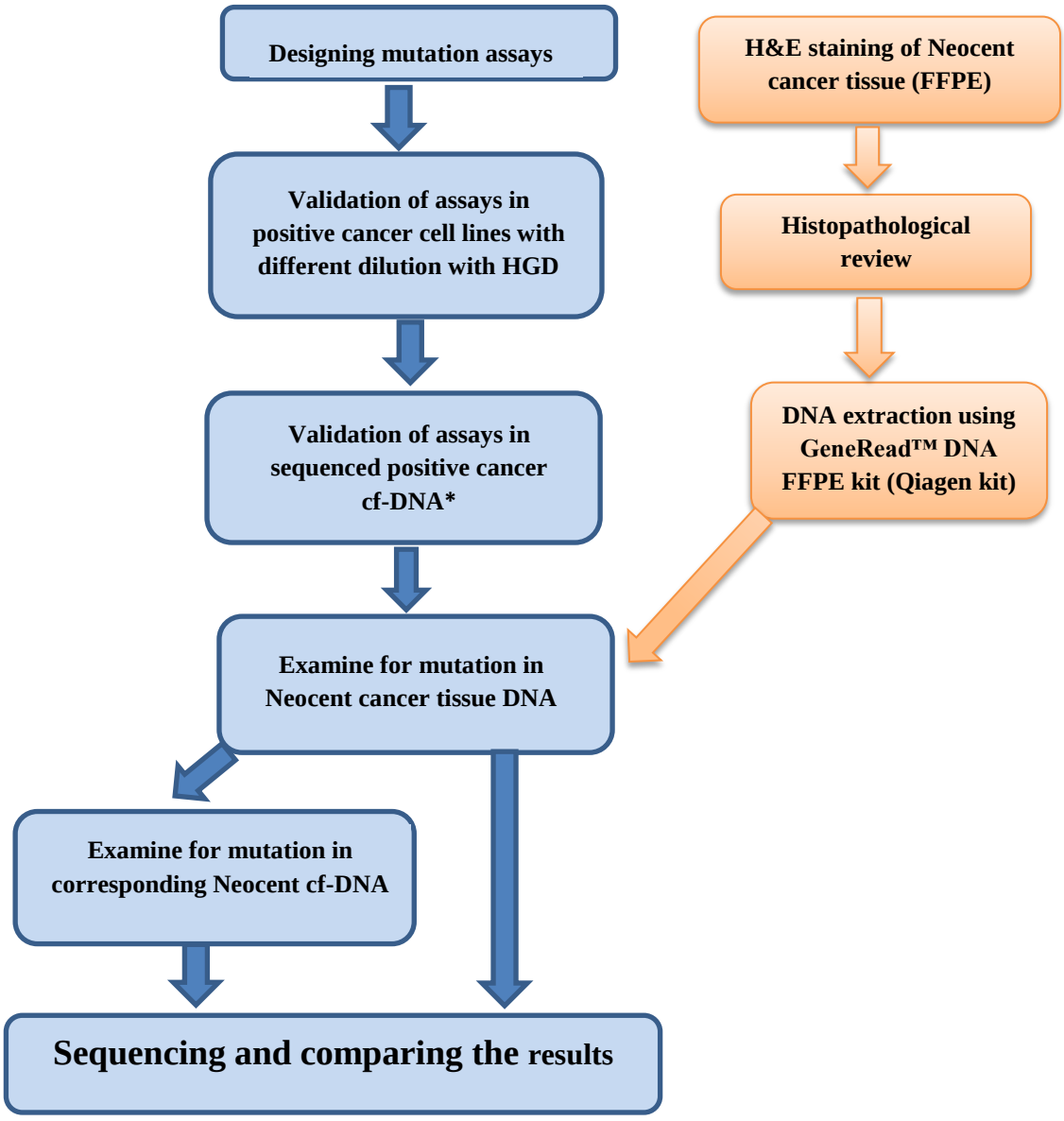


Figure 2.4. The diagram shows the function of PNA. The diagram shows mutant and wild-type amplicon amplification with a forward primer (FP), reverse primer (RP), wild-type probe (WT pb), mutant probe (MUT pb) and PNA. (A) The PNA block the wild-type allele and prevent DNA amplification. (B) PNA didn't block the mutant allele and DNA is amplified.

The *PIK3CA* mutation assays were validated first in positive cancer cell lines (with different dilution with HGD) and second in the available sequenced positive cf-DNA, then used to examine the presence of *PIK3CA* mutation in Neocent breast cancer tissue DNA and the corresponding cf-DNA (Figure 2.5).



*If it is available

Figure 2.5. Workflow of the steps to detect somatic mutation.

All the designed primers and probes were checked for SNPs using COSMIC and UCSC Genome Browser websites, checked for pseudogenes (PubMed NCBI) and primers secondary structures in OligoEvaluator™ (from SIGMA- ALDRICH). The Ensembl Genome Browser website was used to check exons and intron of the genes.

2.2.7.1 Designing of the mutation assays (primers, probes and PNAs)

The *PIK3CA* and *TP53* mutation assays (Forward, reverse primers and specific allele probes) were designed using Primer 3 and Primer Express. The specific allele probes (the FAM and VIC were used as a reporter for the mutant allele and wild-type allele respectively, and NFQ-MGB as Quencher) were used to detect *PIK3CA* mutations or *TP53* mutations (Table 2.13, 2.14, 2.15, 2.16, 2.17 & 2.18). Reverse primers, which were used to detect *PIK3CA* c.1633G>A and c.1624G>A mutations was ligated to LNA (Locked Nucleic Acid) at 3' (in the last base) to avoid the amplification of a pseudogene on chromosome 22 that has >95% homology to exon 9 of *PIK3CA*. LNA is a nucleic acid analog that contains a methylene bridge between 2'-O,4'-C in the ribose structure (Pfundheller et al., 2005). A PNA was used to block the amplification of the wild-type allele. The LNA was used to increase the specificity, while the PNA was used to raise the sensitivity of the PCR reaction. The stock concentrations were 100 µM.

***PIK3CA* c.1624G>A**	**chr3:179,218,242-179,218,354**		**Amplicon size =113 bp**
	Sequence	Tm °C	% GC	Position
Forward primer	5' GGAAAATGACAAAGAACAGCTCAA 3'	58.5	38	Sense
Wild-type probe (VIC/MGB)	5' CTCTCTCTGAAATCAC 3'	68.0	44	Sense
Mutant probe (FAM/MGB)	5' CTCTCTCTAAAATCAC 3'	66.0	38	Sense
Reverse Primer +LNA	5' ATCTCCATTTTAGCACTTACCTGTG{A} 3'	58.4	38	Antisense
PNA (Purines%= 33.3%)	5' CTCTCTCTGAAATCACTG 3'	69.7	44.4	Sense

Table 2.13. The *PIK3CA* (c.1624G>A) mutation assay sequences. The *PIK3CA* (c.1624G>A) mutation assay sequence, Annealing temperature, GC content (%) and position of the primers, allele specific probes and PNA. The LNA place at last base (between brackets) of the reverse primer. The blue coloured base is the wild-type allele, and the red coloured base is the mutant allele.

***PIK3CA* c.1633G>A**	**chr3:179,218,242-179,218,354**		**Amplicon size=113 bp**
	Sequence	Tm °C	% GC	Position
Forward primer	5' GGAAAATGACAAAGAACAGCTCAA 3'	58.5	38	Sense
Wild-type probe (MGB/VIC)	5' ATCACTGAGCAGGAGAA 3'	69.0	47	Sense

Mutant probe (MGB\FAM)	5’ ATCACTAAGCAGGAGAA 3’	66.0	41	Sense
Reverse Primer +LNA	5’ ATCTCCATTTTAGCACTTACCTGTG{A} 3’	58.4	38	Antisense
PNA (Purines%= 58.3%)	5’ TCACTGAGCAGG 3’	70.8	58.3	Sense

Table 2.14. The *PIK3CA* (c.1633G>A) mutation assay sequences. The *PIK3CA* (c.1633G>A) mutation assay sequence, Annealing temperature, GC content (%) and position of the primers, allele specific probes and PNA. The LNA place at last base (between brackets) of the reverse primer. The blue coloured base is the wild-type allele, and the red coloured base is the mutant allele.

***PIK3CA* c.3140A>G**	**chr3:179,234,254-179,234,414**	**Amplicon size=161 bp**		
	Sequence	Tm °C	% GC	Position
Forward primer	5’ CAAGAGGCTTTGGAGTATTTCATG 3’	58	42	Sense
Wild-type probe (MGB\VIC)	5’ ATGCACATCATGGTGG-3’	67	50	Sense
Mutant probe (MGB\FAM)	5’ ATGCACGTCATGGTG 3’	66	53	Sense
Reverse Primer	5’ ACAGTGCAGTGTGGAATCCAGA 3’	59.2	50	Antisense
PNA (Purines%= 53.3%)	5’ ATGCACATCATGGTG 3’	74	46.7	Sense

Table 2.15. The *PIK3CA* (c.3140A>G) mutation assay sequences. The *PIK3CA* (c.3140A>G) mutation assay sequence, Annealing temperature, GC content (%) and position of the primers, alle specific probes and PNA. The blue coloured base is wild-type allele, and the red coloured base is the mutant allele.

***TP53* (c.818G>A)**	**chr17:7673737-7673835**	**Amplicon size =98 bp**		
	Sequence	Tm °C	% GC	Position
Forward primer	5’ GTAATCTACTGGGACGGAACAGCT 3’	58.8	50	Sense
Reverse primer	5’ GCTCCCCTTTCTTGCGGA 3’	59.9	61	Anti-sense
General probe (FAM/MGB)	5’ CACAGAGGAAGAGAATC 3’	70	47	Sense
Wild-type probe (VIC/MGB)	5’ TGAGGTGCGTGTTTGT 3’	69	50	Sense
Mutant probe (FAM/MGB)	5’ AGGTGCATGTTTGTGC 3’	68	50	Sense
PNA (Purines%= 53.8%)	5’ AGGTGCGTGTTTG 3’	72.9	53.8	Sense

Table 2.16. The *TP53* (c.818G>A) mutation assay sequences. The *TP53* (c.818G>A) mutation assay sequence, Annealing temperature, GC content (%) and position of the primers, general probe, allele-specific probes and PNA. The blue coloured base is wild-type allele, and the red coloured base is the mutant allele.

***TP53*-(c.524G>A)**	**chr17:7,675,019-7,675,131**	**Amplicon size= 112 bp**		
	Sequence	Tm °C	% GC	Position
Forward primer	5' GCCATCTACAAGCAGTCACAGC 3'	58.6	55	Sense
Reverse primer	5' CAGCCCTGTCGTCTCTCCAG 3'	59.6	65	Anti-sense
General probe (FAM/MGB)	5' CAGATAGCGATGGTGAGC 3'	69	56	Sense
Wild-type probe (VIC/MGB)	5' TTGTGAGGCGCTGCC 3'	70	67	Sense
Mutant probe (FAM/MGB)	5' TGAGGCACTGCCC 3'	69	69	Sense
PNA (Purines%= 54.5%)	5' TGAGGCGCTGC 3'	74.3	72.2	Sense

Table 2.17. The *TP53* (c.524G>A) mutation assay sequences. The *TP53* (c.524G>A) mutation assay sequence, Annealing temperature, GC content (%) and position of the primers, general probe, allele-specific probes and PNA. The blue coloured base is wild-type allele, and the red coloured base is the mutant allele.

***TP53* (c.743G>A)**	**chr17:7,674,190-7,674,286**	**Amplicon size =96 bp**		
	Sequence	Tm °C	% GC	Position
Forward primer	5' GCTCTGACTGTACCACCATCCA 3'	58.7	55	Sense
Reverse primer	5' TCCAGTGTGATGATGGTGAGGA 3'	59.9	50	Anti-sense
General probe (FAM/MGB)	5' CATGCAGGAACTGTTAC 3'	69	47	Anti-sense
Wild-type probe (VIC/MGB)	5' CATGAACCGGAGGCC 3'	69	67	Sense
Mutant probe (FAM/MGB)	5' CATGAACCAGAGGCC 3'	69	60	Sense
PNA (Purines%= 35.7%)	5' GGCCTCCGGTTCAT 3'	73.3	64.3	Anti-sense

Table 2.18. The *TP53* (c.743G>A) mutation assay sequences. The *TP53* (c.743G>A) mutation assay sequence, Annealing temperature, GC content (%) and position of the primers, general probe, allele-specific probes and PNA. The blue coloured base is wild-type allele, and the red coloured base is the mutant allele.

2.2.7.2 Validation of *PIK3CA* and *TP53* mutation assays using qPCR

Validation of the *PIK3CA* mutation was performed using a peptide nucleic acid (PNA) clamp/TaqMan probe approach as described previously (Rosell et al., 2009) on a StepOne Plus thermal cycler.

The validation of PNA sensitivity and specificity in blocking the wild-type allele was carried out. Also, to examine if PNA could increase the sensitivity of mutant probe in detecting low mutant fraction in the presence of a large background of wild-type alleles. HGD (10 ng/3.6 µL) was used as a negative control of the presence of mutant allele. DNAs isolated from 3 different cell lines, HCT-116, MCF7 and SW948, which have 3 different *PIK3CA* mutations (c.3140A>G, c.1633G>A and c.1624G>A) were used as positive control and their serial dilution with HGD (10 ng/3.6 µL). The PCR products were amplified through qPCR (Table 2.19) for 50 cycles, with genotyping mastermix, in duplex (both mutant and wild-type probes) and with and without the PNA.

	Temperature	Time	Cycles
Initial denaturation	95 °C	10 minutes	
Denaturation	95 °C	15 seconds	50 cycles
Annealing/Extension	60 °C	1 minute	

Table 2.19. PCR conditions to amplify specific DNA regions with *PIK3CA* or *TP53* genes.

The stock concentration of primers, probes and PNAs was 100 µM. Forward primers, Reverse primers and PNA were diluted to 1:10 (10 µM) with ultra-pure water, while probes were diluted to 1:50 (2 µM). 10 µL of PCR reaction volume was loaded in duplicate/triplicate for each sample in 96-well plates. The Applied Biosystems StepOne Plus™ Real-Time PCR System Thermal Cycling Block machine with TaqMan genotyping mastermix reagent was used to detect *PIK3CA* mutations.

The *PIK3CA* gene mutation assays were validated in cf-DNAs which were extracted from the plasma of a patient with metastatic breast cancer and confirmed to have *PIK3CA* mutations by next-generation sequencing by another research group. Also, 10 Neocent tissues DNA, their matched plasma DNA and lymphocyte DNA were used as a pilot study to examine for the presence of *PIK3CA* mutation. The 10 Neocent tissue DNAs and their matched plasma and lymphocyte DNAs were pre-amplified (around 300-400 ng/ µL) using TaqMan Pre-Amp mastermix and 10x Pre-Amp primer mix (*PIK3CA* c.3140A>G, c.1624G>A and c.1633G>A forward and reverse primers) in the

Veriti thermal cycler (Table 2.20). After that, the amplified DNA was diluted 1:10 with sterile water and used to validate *PIK3CA* mutation assays.

	Temperature and duration	Cycles
Initial denaturation	95 °C for 10 minutes	1 cycle
Denaturation	95 °C for 15 seconds	10 cycles
Annealing	60 °C for 4 minutes	
Extension	99 °C for 10 minutes	1 cycle

Table 2.20. Thermal cycling conditions used in amplifying tissue and plasma Neocent DNAs.

The peptide nucleic acids (PNAs) for the three different *TP53* gene mutations were designed to block the wild-type alleles. The PNAs specificity in blocking wild-type alleles was validated in homozygous cell line DNAs (SW948) and HGD, through qPCR for 50 cycles and using general probes and genotyping mastermix.

2.2.7.3 Droplet Digital PCR (ddPCR)

Validation of the *PIK3CA* and *TP53* mutations assays was performed using a Bio-Rad QX200 digital droplet PCR system as described previously (Hindson et al., 2011).

The *PIK3CA* and *TP53* mutations assays [forward primers, reverse primers and allele-specific probes (FAM/MGB mutant probe and VIC/MGB Wild-type probe)] (Table 2.13, 2.14, 2.15, 2.16, 2.17 & 2.18) were validated using the ddPCR method. The sensitivity of the assay was determined by generating serial dilutions of the cell line DNA (mutant allele) with wild-type DNA (HGD). The amplicons, which contain the hotspot mutation area were pre-amplified in HGD, cell lines DNA and their serial dilution with HGD. The pre-amplified and matched amplified DNA was used as a template to validate the ddPCR method in detection of the mutant allele and compare the efficiency of mutant allele detection between pre-amplified and non-amplified DNA. The pre-amplification method was validated to be used with plasma DNA due to its very low concentration. After the validation of ddPCR in detection of the mutant allele in a serial dilution of cell line DNA, 26 Neocent tissue DNA samples at baseline (T1) were examined to detect the presence of *PIK3CA* mutations or *TP53* mutations. Followed by examination of the matched Neocent tissue DNA after completing treatment (T3) for the positive baseline (T1) samples.

The 10x Pre-Amp primer mix consisted of a primer pool of *PIK3CA* (c.3140A>G), (c.1633G>A), (c.1624G>A) and *TP53* (c.524G>A), (c.743G>A) and (c.818G>A) forward and reverse primers (Table 2.21). TaqMan Pre-Amp mastermix and 10x Pre-Amp primer mix were used in the pre-amplification reaction (Table 2.22) with HGD and cell line DNA (10 ng/3.6 µL), then amplified for five cycles using Veriti thermal cycler (Table 2.20).

Reagent	Volume	Stock concentration
Each Forward or reverse primers	1.5 µL	100 µM
Sterile water	Up to 500 µL	-

Table 2.21. The components of PreAmp primer mix. The reagents and volume of PreAmp primer mix.

Component	Volume
Taq Man Pre-Amp mstermix	12.5 µL
10x Pre-Amp primer mix	2.5 µL
DNA	10 ng/3.6 µL
Sterile water	6.4 µL
Total	25 µL

Table 2.22. The pre-amplification reaction. The components and the volume of the pre-amplification reaction.

The mastermix for 20x target (mutant allele) and 20x reference (wild-type allele) were prepared from forward, reverse primers and probe (VIC/MGB labelled probe for the reference, and FAM/MGB labelled probe for the target) (Table 2.23).

Components	Volume
Forward primer (200 µM stock)	2.25 µL
Reverse primer (200 µM stock)	2.25 µL
Probe (100 µM stock)	2.5 µL
Sterile, ultra-pure water	43 µL
Total	50 µL

Table 2.23. The components of ddPCR mastermix. The element and volume of ddPCR master mix.

The reaction volume composed of ddPCR supermix, reference gene mastermix, target gene mastermix and DNA was prepared (Table 2.24). A 20 µL the reaction volume was loaded into the DG8 cartridge with 70 µL of droplet generation oil to generate the

droplets in the QX200 droplet generator, followed by qPCR for 40 cycles with annealing temperature = 59 °C (Table 2.25).

Reagents	Volume
2x ddPCR supermix for probes	11 µL
20X Reference (VIC labelled probe) /forward and reverse primers	1.1 µL
20X Target (FAM labelled probe) /forward and reverse primers	1.1 µL
DNA	10 ng/3.6 µL
Sterile water	To 22 µL

Table 2.24. The component and volume of the ddPCR reaction.

Stages	Temperature	Time	Cycles
Initial denaturation	95 °C	10 minutes	
Denaturation	94 °C	30 seconds	40 cycles
Annealing/Extention	59 °C	1 minute	
Final extention	98 °C	10 minutes	

Table 2.25. The PCR stages of the ddPCR.

The ddPCR results were read with the QX200 droplet reader and analysed using Quantasoft software. The thresholds were set manually for wild-type and mutant droplets based on the results of HGD (wild-type allele) and positive cell line DNA (mutant allele) (Figure 2.6).

The *PIK3CA* CNA assays (Forward, reverse primers and FAM/MGB probes) were designed used Primer 3 and Primer Express programmes (Table 2.26). The *RPPH1* gene assay (Forward, reverse primers and VIC/MGB probes) was used as the reference to detect *PIK3CA* CNA using ddPCR (Table 2.27). The *RPPH1* gene was chosen as a reference gene as it doesn't consist of amplification (Shaw et al., 2012).

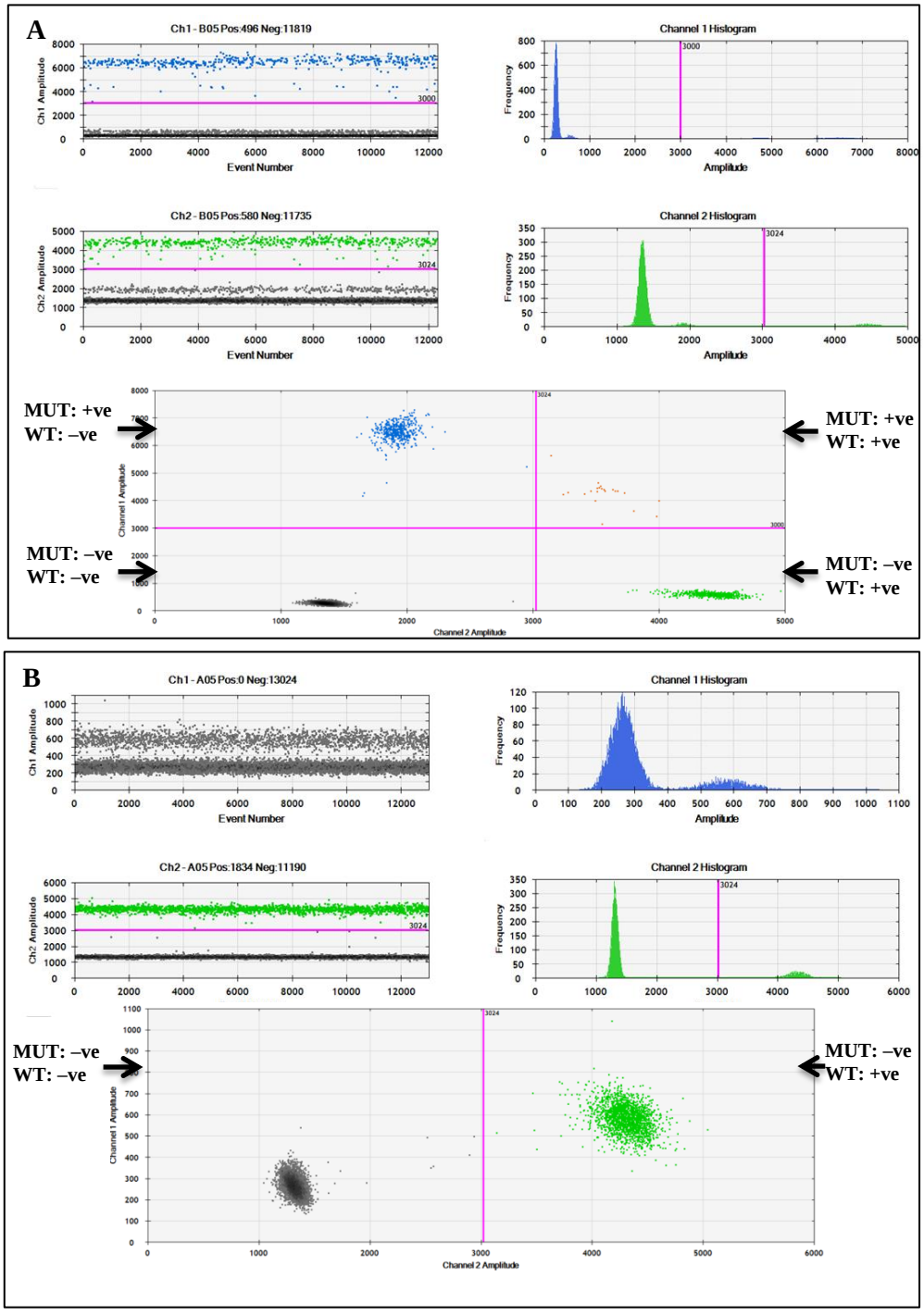


Figure 2.6. The ddPCR results. (A) The ddPCR results of *TP53* (c.524 G>A) in CCRF-CEM cell line DNA. (B) The ddPCR results of *TP53* (c.524 G>A) in HGD. The thresholds were set manually for wild-type and mutant droplets based on the results of HGD (wild-type allele) and positive cell line DNA (mutant allele). The blue dots represent the droplets with mutant allele (MUT), the green dots represent the droplets with wild-type allele (WT), and the black dots represent the droplets with no template.

Gene	179,204510-179,204590	Tm °C	% GC	Position
***PIK3CA* CNAs-FP**	5' GTTCGAACAGGTATCTACCATGGA 3'	58.1	46	Sense
***PIK3CA* CNAs-RP**	5' CTGGGATTGGAACAAGGTACTCTT 3'	58.6	46	Anti-sense
***PIK3CA* CNAs-Probe (FAM/MGB)**	5' AACCCTTATGTGACAATGT 3'	69	37	Sense

Table 2.26. The *PIK3CA* CNA assay sequences. FP; Forward primer, RP; Reverse primer.

Gene	Sequence	Tm °C	% GC	Position
***RPPH1*-FP**	5' CGGAGGGAAGCTCATCAGTG 3'	59.4	60	Sense
***RPPH1*-RP**	5' GACATGGGAGTGGAGTGACA 3'	55.3	55	Anti-sense
***RPPH1*-Probe (VIC-MGB)**	5' CACGAGCTGAGTGCGT 3'	69	63	Sense

Table 2.27. The *PPPH1* assay sequences. FP; Forward primer, RP; Reverse primer.

The RERF-LC-Sq1 (squamous cell lung cancer) cell line was used as positive control for *PIK3CA* CNA as it has amplification in the DNA region (3:176228330-180305959) (cancer.sanger.ac.uk/COSMIC).

2.2.7.4 High resolution melting curve method (HRMC)

The HRMC method was chosen to detect *TP53* gene mutations due to multiple *TP53* mutations in the same amplicon and GC bases rich area, which make designing for the specific base assays more challenging. The aim of the HRMC method was to detect relative fluorescence signal differences from 60 °C to 95 °C (melt curve) between DNA with the mutant allele and the reference DNA, which consist of wild-type allele (e.g., HGD). As the melting curve of the amplicon with any *TP53* mutant allele should be different than melting curve of the amplicon with *TP53* wild-type allele (HGD), so the HRMC method is not specific to detect specific *TP53* mutation but could detect any *TP53* mutations in the amplicon.

The *TP53* primers and PNAs were validated using the HRMC method using Step One PCR thermal cycling version 2.3 and MeltDoctor mastermix. The *TP53* mutations assays were validated through one step of qPCR with PNA, where the particular mutation area in DNA was amplified for ten cycles of the touch-down method to increase the specificity of the primers and avoid amplifying non-specific sequence, followed by other 30 cycles then a melting curve (Table 2.28). Annealing temperatures

of the primers were 60 °C. The validations were carried out using HGD and diluted positive cell line DNA for *TP53* mutations with HGD.

	Temperature	Time	Cycles
Initial Denaturation	**95** °C	**10 minutes**	
Denaturation	**95** °C	**1 second**	**10 cycle (Touch down)**
Annealing/Extension	**70** °C	**20 seconds**	
Denaturation	**95** °C	**1 second**	**30 cycles**
Annealing/Extension	**60** °C	**20 seconds**	
Melt curve stage	**95** °C	**10 seconds**	
	60 °C	**1 minute**	
	95 °C	**15 seconds**	
	60 °C	**15 seconds**	

Table 2.28. The PCR steps of HRMC. The *TP53* mutations areas in DNA were amplified for ten cycles of the touch-down method, followed by other 30 cycles then a melting curve.

The HRMC results (melt region temperature data and melt region normalised data) were analysed manually using HGD as a reference compared to the other DNA samples: 1-Calculate average of melt region temperature data for each reading for all the samples (59.9 °C – 94 °C) (~430 reading).

2- Using melt region normalised data (Normalised fluorescence signal value) to select the temperatures regions with the florescence signal value changes (Figure 2.7).

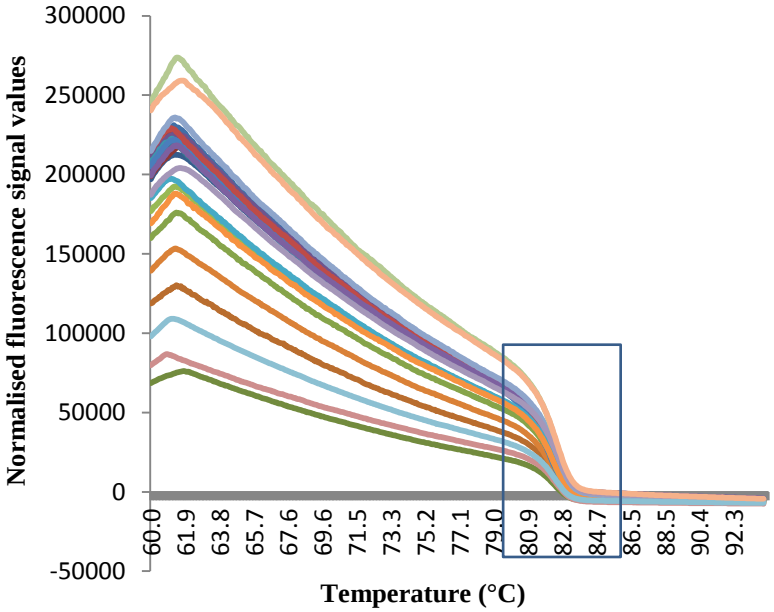


Figure 2.7. Melt region normalised data. The rectangular area showed the selected temperature with normalised fluorescence signal values changes.

3- Calculate the percentage of melt region normalised data (Normalised fluorescence signal value) according to smallest and highest selected melting temperatures +/- 0.5 °C (Figure 2.8).

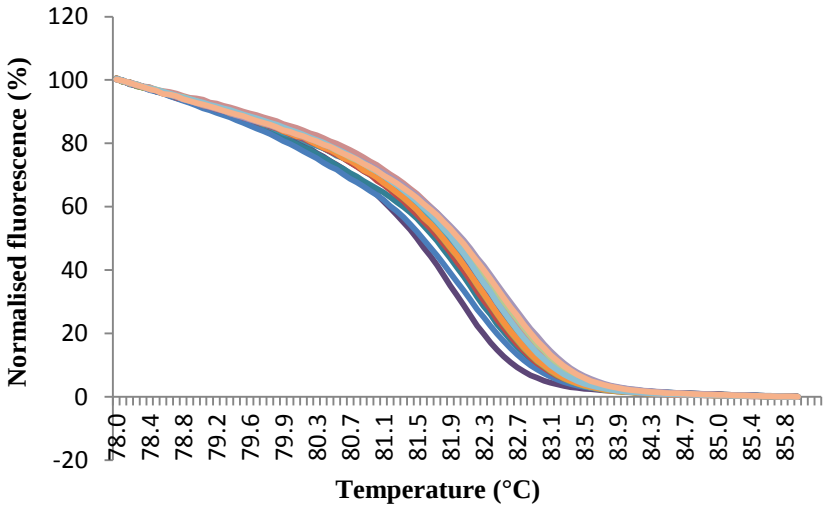


Figure 2.8. The percentage of melt region normalized data. The percentage of normalised fluorescence signal according to the smallest and the highest chose melting temperatures +/- 0.5 °C.

4- Calculate the difference between the normalised fluorescence percentage of DNA sample and the reference to get the melting curve (Figure 2.9).

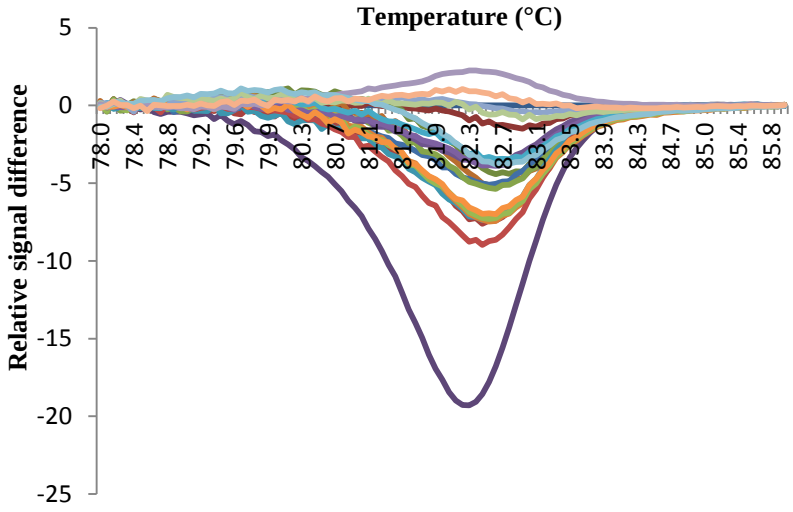


Figure 2.9. The relative fluorescence signal differences (melt curve). The relative normalised fluorescence percentage differences between DNA samples and the reference (HGD).

After validation of the HRMC method in a diluted cell line DNA. The presence of *TP53* gene mutations was examined in 26 Neocent tissue DNAs and their paired lymphocyte DNA, using the HRMC method.

2.2.7.5 DNA Sanger sequencing

DNA Sanger sequencing was used to detect the presence or absence of single-base mutations in *PIK3CA* and *TP53* genes.

2.2.7.5.1 DNA amplification

Amplifying the different amplicons in the three hot spot mutations regions in *PIK3CA* and *TP53* genes in the selected Neocent tissue DNA which were positive for *PIK3CA* mutations by qPCR or ddPCR or positive for *TP53* mutations according to ddPCR or HRMC results. The Platinum Taq DNA high-fidelity polymerase was used to amplify different amplicons from tissue DNA in the Neocent samples (Table 2.29). The Veriti thermal cycling machine was used to amplify each PCR reaction (Table 2.30).

Component	Volume	Final concentration
10X High Fidelity PCR Buffer	5 µL	1X
50 mM $MgSO_4$	2 µL	2.0 mM
10 mM dNTP Mix	1 µL	0.2 mM
10 µM forward primer	1 µL	0.2 µM
10 µM Reverse primer	1 µL	0.2 µM
Template DNA	10 ng/3.6 µL	<500ng
Platinum® Taq DNA Polymerase High Fidelity (5 U/ µL)	0.2 µL	1 U
Autoclaved, distilled water	To 50 µL	-

Table 2.29. The components of PCR reaction. The component and volumes of each PCR reaction to amplify DNA.

Step	Temprature (°C)	Time	Cycles
Initial Denaturation	94	2 minutes	
Denaturation	94	15 seconds	35 cycles
Anneal	58	30 seconds	
Extend	68	30 seconds	

Table 2.30. The PCR conditions. Step, temperatures and times for PCR reaction to amplify DNA, using Veriti thermal cycling.

2.2.7.5.2 DNA purification

The PCR products were purified using the USB® ExoSAP-IT® PCR Product Cleanup kit. Each 5 µL of post-PCR reaction product was added to 2 µL of ExoSAP-IT reagent, then incubated at 37 °C for 15 minutes to degrade remaining primers and nucleotide. After that, it incubated at 80 °C for 15 minutes to inactivate the ExoSAP-IT reagent. DNA (PCR product) was quantified using the Qubit fluorometer.

2.2.7.5.3 DNA Sanger sequencing

DNA Sanger Sequencing of double strand PCR products was carried out using a high-throughput Applied Biosystems 3730 Genetic Analyser in The Protein Nucleic Acid Chemistry Laboratory (PNACL), University of Leicester. The DNA Sanger sequencing required DNA concentration ranged between 1-3 ng/ µL and primer concentration=0.8-1.0 pmol/ µL.

Chapter 3. Circulating micro-RNA expression in breast cancer

3.1 Cancer-specific miRNAs

The hypothesis to be tested is that specific circulating miRNAs may have utility as biomarkers of breast cancer. To achieve this, specific miRNAs are expected to have a significant difference in their expression levels in cancer samples compared to control samples.

To detect differences in miRNAs profiles between cancers and controls, the expression level of 384 miRNAs were examined using TaqMan microRNA Arrays. The miRNAs were extracted from plasma samples of 35 patients diagnosed with breast cancer before starting neoadjuvant treatment (P1) in the Neocent trial. Groups of plasma samples were pooled together according to the type of neoadjuvant treatment received (chemotherapy and endocrine therapy) and the response to the treatment, based on radiological response after neoadjuvant treatment (Table 3.1). The pooled cancer samples were compared with pooled plasma miRNA samples from healthy female controls.

The expression levels of 384 miRNAs were compared in the pooled cancer sample and individual cancer samples to validate the pooling method of different cancer samples together. The single cancer sample was selected randomly from each group. Two different groups of control samples [Leicester controls (n=5) and London controls (n=12)] were pooled and used as control groups to compare their microRNA expression to the pooled and single cancer samples.

Time of sample	Name of group	Number of pooled samples
Baseline (P1)	Response to chemotherapy (RC)P1	n=9
	No response to chemotherapy (NRC)P1	n=6
	Response to endocrine therapy (RE)P1	n=13
	No response to endocrine therapy (NRE)P1	n=7
After completing treatment (P3)	Response to chemotherapy (RC)P3	n=9
	No response to chemotherapy (NRC)P3	n=6
	Response to endocrine therapy (RE)P3	n=13
	No response to endocrine therapy (NRE)P3	n=7

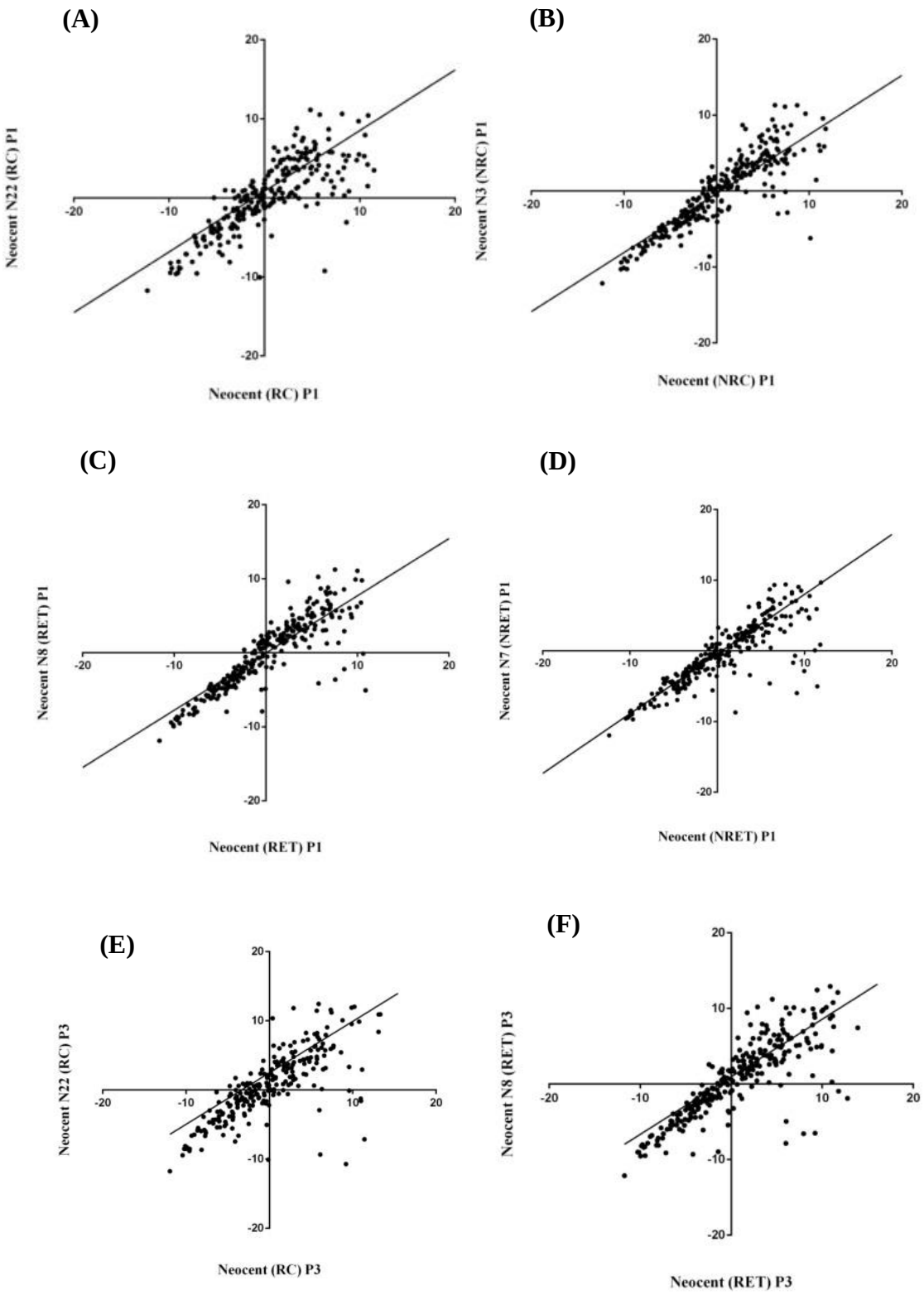
Table 3.1. Pooled samples of breast cancer patient from Neocent study. The cancer samples were divided into eight groups according to the type of treatment received (chemotherapy and endocrine therapy) and their response to the treatment (based on radiological results).

3.1.1 Correlation between pooled and individual cancer samples

There was a significant positive correlation between the expression level of 384 miRNAs in pooled samples and corresponding single samples, chosen randomly from each group (Figure 3.1; Table 3.2). These highly correlated results suggest that the expression level of miRNAs in pooled samples will reflect the expression level of miRNAs in each sample in this pool. Despite strong correlations between pooled and single samples, there were differences in expression profile of a significant number of miRNAs (more than 100 miRNAs). These miRNA expression differences between pooled and single samples could be due to the effect of the high or low expression levels of these miRNAs in an individual sample, compared to the average in the pool.

Correlation (ΔCt)	Spearman (r)	P-value (two-tailed)	95% confidence interval
Pool (RC) P1 & N22 (RC) P1	0.80	< 0.0001	0.76 to 0.83
Pool (NRC) P1 & N3 (NRC) P1	0.87	< 0.0001	0.84 to 0.89
Pool (RET) P1& N8 (RET) P1	0.87	< 0.0001	0.85 to 0.90
Pool (NRET) P1 & N7 (NRET) P1	0.91	< 0.0001	0.89 to 0.93
Pool (RC) P3 & N22 (RC) P3	0.80	< 0.0001	0.76 to 0.83
Pool (RET) P3 & N8 (RET) P3	0.83	< 0.0001	0.79 to 0.86
Pool (NRET) P3 & N7 (NRET) P3	0.92	< 0.0001	0.90 to 0.93

Table 3.2. Correlation analysis between cancer individual samples and pools. The Correlation (Spearman test) between ΔCts of pooled cancer plasma samples and their randomly chosen individual cancer sample, from TaqMan microRNA Array data. ΔCts were calculated in comparing to pooled control samples. RC; chemotherapy responder group, NRC; chemotherapy non-responder group, RET; endocrine therapy responder group, NRET; endocrine therapy non-responder group, P1; plasma samples which were taken before neoadjuvant treatment and P3; plasma samples after neoadjuvant treatment.


(A)
Neocent N22 (RC) P1
Neocent (RC) P1
(B)
Neocent N3 (NRC) P1
Neocent (NRC) P1
(C)
Neocent N8 (RET) P1
Neocent (RET) P1
(D)
Neocent N7 (NRET) P1
Neocent (NRET) P1
(E)
Neocent N22 (RC) P3
Neocent (RC) P3
(F)
Neocent N8 (RET) P3
Neocent (RET) P3

(G)

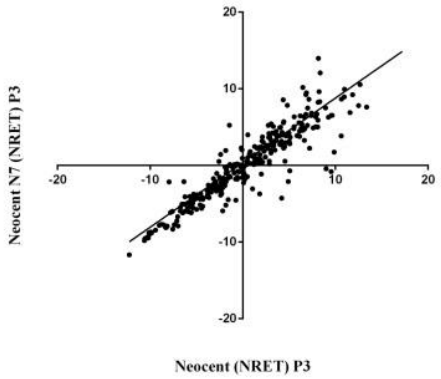


Figure 3.1. Correlation analysis between individual cancer plasma samples and pools. Correlation between pooled and individual cancer samples at baseline (P1), based on the expression results of 384 microRNA from TaqMan microRNA array card. Their ΔCts were calculated compared to the control pooled samples. (A) The correlation between Neocent plasma samples at baseline and response to chemotherapy (Neocent (RC) P1) and their randomly chosen individual sample (N22P1). (B) The correlation between Neocent plasma samples at baseline and doesn't response to chemotherapy (Neocent (NRC) P1) and their randomly selected single sample (N3P1). (C) The correlation between baseline Neocent plasma samples and response to endocrine therapy (Neocent (RET) P1) and their randomly selected single sample (N8P1). (D) The correlation between baseline Neocent plasma samples and doesn't response to endocrine therapy (Neocent (NRET) P1) and their randomly selected single sample (N7P1). (E) The correlation between Neocent plasma samples after neoadjuvant treatment and response to chemotherapy (Neocent (RC) P3) and their randomly selected single sample (N22P3). (F) The correlation between Neocent plasma samples after treatment and response to endocrine therapy (Neocent (RET) P3) and their randomly selected single sample (N8P3). (G) The correlation between Neocent plasma samples after treatment and doesn't response to chemotherapy (Neocent (NRET) P3) and their randomly selected single sample (N7P3). There are 147 data points are outside the axis in graph (A), 125 in graph (B), 127 data points are outside the axis in graph (C), 121 in graph (D), 144 in graph (E), 177 in graph (F) and 102 in graph (G).

3.1.2 Discovery analysis for cancer-specific microRNAs

The expression levels of 384 microRNAs in the pooled cancer baseline sample (P1) and the controls from the TaqMan microRNA Array data were compared to discover cancer-specific miRNAs with significant expression difference, which may serve as screening biomarkers of breast cancer. Two different analysis methods (multiple t-test one per row and MeV software) were used to compare the expression levels of 384 microRNAs in the pooled cancer samples and the controls. The fold change was also calculated for comparison with independent results, generated by Dr D. Guttery in the group. A significant fold-change was defined as > 2.5. The results showed a significant difference in 69 miRNAs expression profiles in cancer samples compared to the controls with *P*-value < 0.05 and fold-change > 2.5, (Figure 3.2; Table 3.3). While according to the MeV analysis, there was a significant difference in 87 miRNAs expression profile in cancer samples compared to the controls, (Figure 3.3 & 3.4; Table

3.4). Both analyses showed an overlap for 49 miRNAs (Figure 3.5), all of which were up-regulated in cancer samples, except *miR-509-5p*, which was down-regulated, compared to control samples.

MicroRNA	*P*-value	Fold change
hsa-miR-598*	0.0001	3.66
hsa-miR-30b*	0.0001	4.02
hsa-miR-652*	0.0005	6.31
hsa-miR-15b	0.0014	5.37
hsa-miR-509-5p	0.0018	-8.29
hsa-miR-671-3p*	0.0025	11.21
hsa-miR-199a-5p	0.0027	10.54
hsa-miR-27a*	0.0027	7.89
hsa-miR-579*	0.0032	9.52
hsa-miR-24	0.0032	3.39
hsa-miR-28-3p	0.0034	4.02
hsa-miR-576-3p	0.0035	9.35
hsa-miR-301b	0.0038	10.28
hsa-miR-889	0.0040	8.83
hsa-miR-142-3p	0.0043	6.54
hsa-miR-655	0.0043	8.43
hsa-miR-758	0.0045	8.60
hsa-miR-708	0.0046	5.46
hsa-miR-362-3p	0.0047	8.00
hsa-let-7a*	0.0058	11.79

Table 3.3. The miRNAs with significant expression level change between cancer and control samples. Top 20 miRNAs with significant expression level change between cancer and control samples. The Comparison was between the expression level of 384 microRNAs from pooled cancer patients at baseline (P1) and the controls, using TaqMan Array data. Top 20 miRNAs which showed different expression profiles (based on ΔCt values) between pooled controls and baseline cancer samples (P1) by Multiple t-test one per row with a significance of *P*-value > 0.05 and fold-change (-ΔΔCt) >2.5. All miRNAs were up-regulated in baseline cancer samples, except *miR-509-5p*, which was down-regulated (red colour). The miRNAs* were chosen for further investigation.

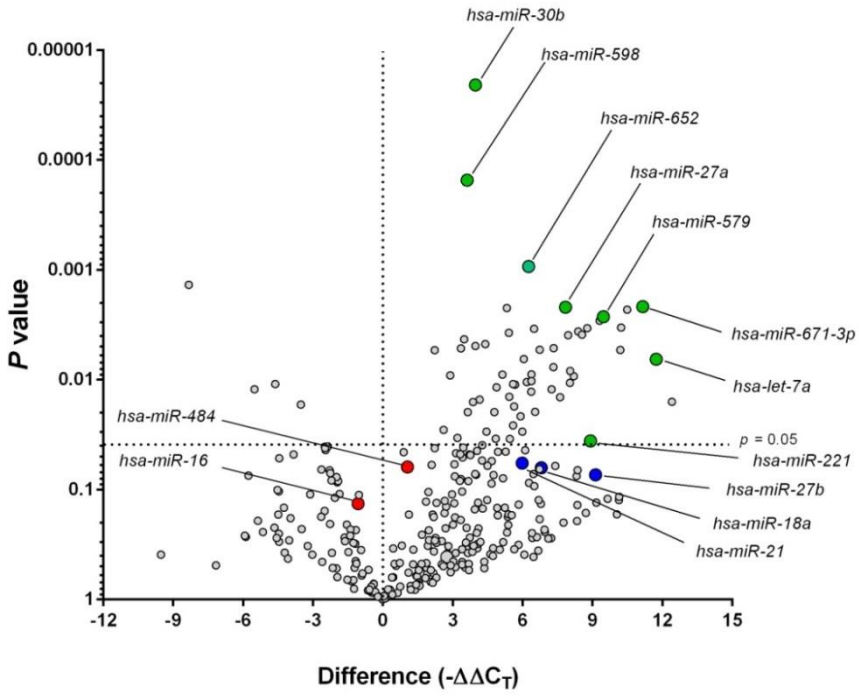


Figure 3.2. Comparison of microRNA expression results between pooled cancer samples and control samples. Volcano plots represent the comparison of expression results of 384 microRNAs from pooled cancer samples and control samples, based on fold-changes (-ΔΔCt) and *P*-values. Green and blue dots represent selected miRNAs for further investigation, and red dots represent two miRNAs which were used as endogenous controls in further investigation.

The miRNAs with significant expression level change between cancer and control samples, according to MeV software results	
Down-regulated miRNA	*hsa-miR-509-5p.*
Up-regulated miRNAs	*hsa-let-7a*, hsa-miR-671-3p*, hsa-miR-579*, hsa-miR-221*, hsa-miR-27a*, hsa-miR-27b*, hsa-miR-652*, hsa-miR-18a*, hsa-miR-21*, hsa-miR-15b , hsa-miR-382, hsa-miR-199a-5p, hsa-miR-301b, hsa-miR-23b , hsa-miR-576-3p, hsa-miR-889, hsa-miR-758, hsa-miR-655, hsa-miR-362-3p, hsa-miR-324-5p, hsa-miR-127-3p, hsa-miR-146b-3p, hsa-miR-128a, hsa-miR-152, hsa-miR-744, hsa-miR-654-3p, hsa-miR-331-5p, hsa-miR-34a, hsa-miR-22, hsa-miR-886-3p, hsa-miR-107, hsa-miR-199b-5p, hsa-miR-329, hsa-miR-148b, hsa-let-7f, hsa-miR-142-5p, hsa-miR-335, hsa-miR-495, hsa-miR-145, hsa-miR-142-3p, hsa-miR-338-3p, hsa-miR-551b, hsa-miR-409-5p, hsa-miR-381, hsa-miR-224, hsa-miR-625, hsa-miR-337-5p, hsa-miR-597, hsa-miR-326, hsa-miR-450a, hsa-miR-328, hsa-miR-548d-3p, hsa-miR-429, hsa-miR-9, hsa-miR-379, hsa-miR-708, hsa-miR-888, hsa-miR-199a-3p, hsa-miR-361-5p, hsa-miR-370, hsa-miR-654-5p, hsa-let-7c, hsa-miR-101, hsa-miR-138, hsa-miR-132, hsa-miR-485-3p, hsa-miR-485-3p, hsa-miR-148a, hsa-miR-548d-5p, hsa-miR-487b, hsa-miR-494, hsa-miR-98, hsa-miR-26a, hsa-miR-23a, hsa-miR-106b, hsa-miR-28-5p, hsa-miR-130b, hsa-miR-30c, hsa-miR-130a, hsa-miR-125b, hsa-miR-627, hsa-miR-20a, hsa-miR-29b, hsa-miR-103, hsa-miR-25, hsa-miR-26b, hsa-miR-340.*

Table 3.4. The miRNAs with significant expression level change between cancer and control samples with MeV analysis. The comparison was between the expression level of 384 microRNAs from pooled cancer patient at baseline (P1) and the controls, using TaqMan Array data and MeV analysis. All miRNAs were up-regulated in cancer plasma samples, except *miR-509-5p*, which was down-regulated, compared to the control plasma samples. The miRNAs* were chosen for further investigation.

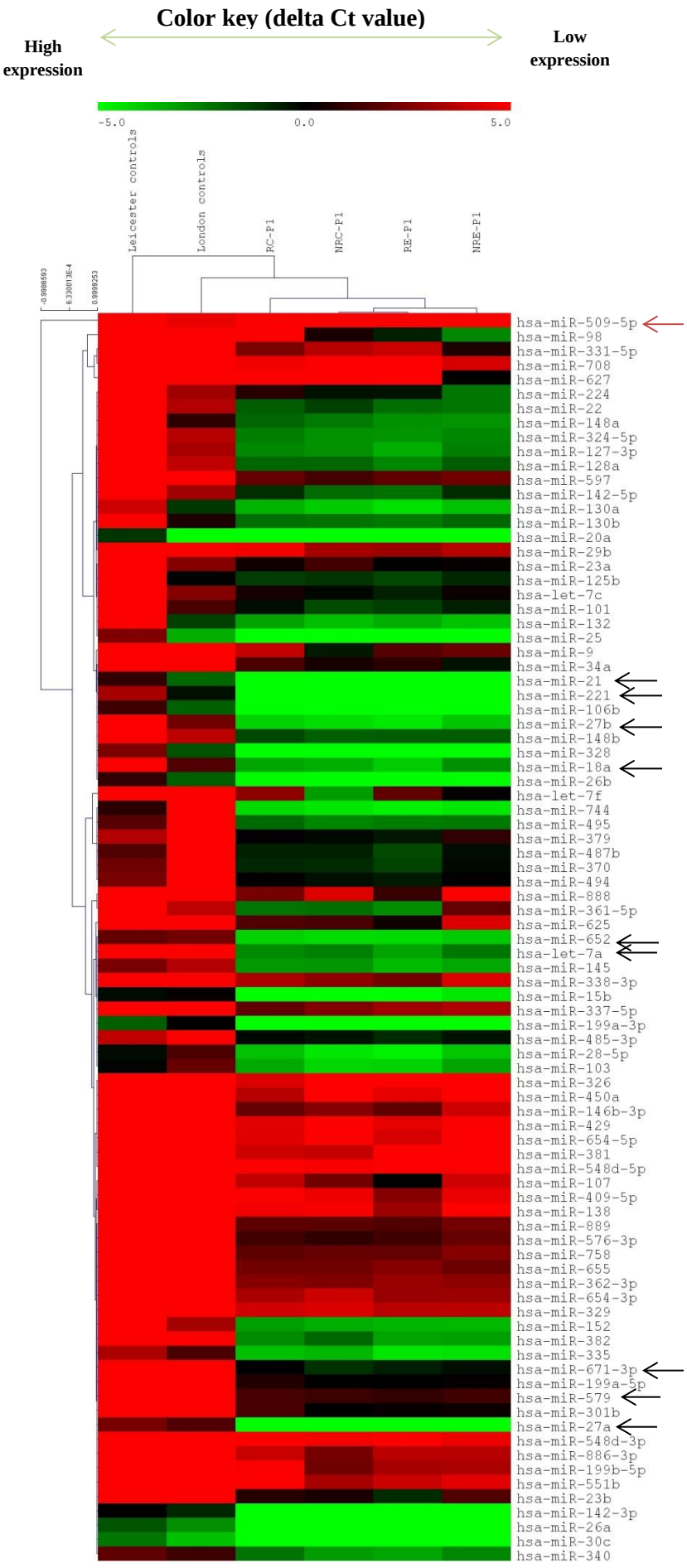

Color key (delta Ct value)
High expression
Low expression
-5.0
0.0
5.0
Leicester controls
London controls
RC-P1
NRC-P1
RE-P1
NRE-P1
hsa-miR-509-5p
hsa-miR-98
hsa-miR-331-5p
hsa-miR-708
hsa-miR-627
hsa-miR-224
hsa-miR-22
hsa-miR-148a
hsa-miR-324-5p
hsa-miR-127-3p
hsa-miR-128a
hsa-miR-597
hsa-miR-142-5p
hsa-miR-130a
hsa-miR-130b
hsa-miR-20a
hsa-miR-29b
hsa-miR-23a
hsa-miR-125b
hsa-let-7c
hsa-miR-101
hsa-miR-132
hsa-miR-25
hsa-miR-9
hsa-miR-34a
hsa-miR-21
hsa-miR-221
hsa-miR-106b
hsa-miR-27b
hsa-miR-148b
hsa-miR-328
hsa-miR-18a
hsa-miR-26b
hsa-let-7f
hsa-miR-744
hsa-miR-495
hsa-miR-379
hsa-miR-487b
hsa-miR-370
hsa-miR-494
hsa-miR-888
hsa-miR-361-5p
hsa-miR-625
hsa-miR-652
hsa-let-7a
hsa-miR-145
hsa-miR-338-3p
hsa-miR-15b
hsa-miR-337-5p
hsa-miR-199a-3p
hsa-miR-485-3p
hsa-miR-28-5p
hsa-miR-103
hsa-miR-326
hsa-miR-450a
hsa-miR-146b-3p
hsa-miR-429
hsa-miR-654-5p
hsa-miR-381
hsa-miR-548d-5p
hsa-miR-107
hsa-miR-409-5p
hsa-miR-138
hsa-miR-889
hsa-miR-576-3p
hsa-miR-758
hsa-miR-655
hsa-miR-362-3p
hsa-miR-654-3p
hsa-miR-329
hsa-miR-152
hsa-miR-382
hsa-miR-335
hsa-miR-671-3p
hsa-miR-199a-5p
hsa-miR-579
hsa-miR-301b
hsa-miR-27a
hsa-miR-548d-3p
hsa-miR-886-3p
hsa-miR-199b-5p
hsa-miR-551b
hsa-miR-23b
hsa-miR-142-3p
hsa-miR-26a
hsa-miR-30c
hsa-miR-340

Figure 3.3. The heat map of Euclidean hierarchical complete linkage clustering of the miRNAs with significant expression level change between cancer and control samples, according to MeV software results. The miRNA expression profiles were compared in six different groups; Leicester controls, London controls, response to chemotherapy (RC-P1), non-response to chemotherapy (NRC-P1), response to endocrine therapy (RE-P1) and non-response to endocrine therapy (NRE-P1). Pointed miRNAs with black arrows were chosen for further investigation. A red arrow pointed to down-regulated miRNA (*miR-509-5p)*.

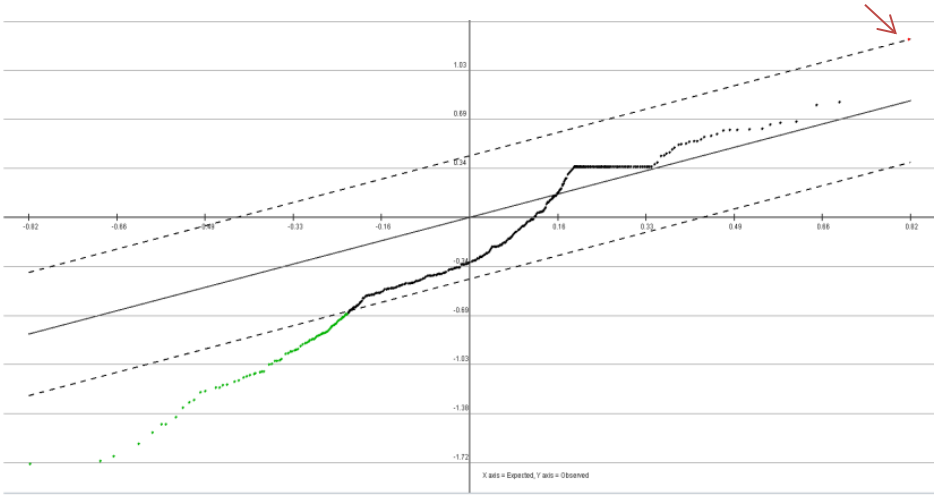


Figure 3.4. Comparison of microRNAs expression results between pooled cancer samples and control samples, according to MeV software results. The comparison was between the expression level of 384 microRNAs from pooled cancer patient at baseline (P1) and the controls, using TaqMan Array data. Green dots represent 86 up-regulated miRNAs, red dot which pointed by the arrow represent the down-regulated miRNA, and the black dots represents miRNAs which showed the insignificant difference in their expression level between cancer and control samples. X axis =expected values and Y axis= observed values.

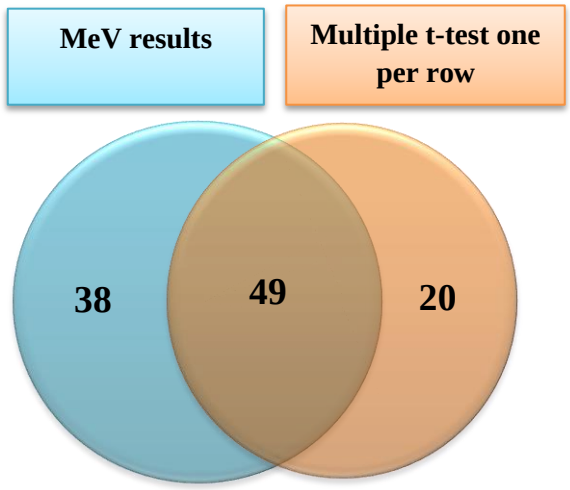


Figure 3.5. The number of microRNAs with significant expression results according to MeV software analysis and multiple t-test one per row (*P*<0.05).

These results were supported by independent findings from Dr D. Guttery, who carried out the same experiment independently (from plasma extraction through to TaqMan Array card and Dunn's multiple comparison test was used for data analysis) and found an overlap with these results (Table 3.5) (Palmieri et al., 2014).

miRNAs	Dr D. Guttery results		Results for this thesis		
	***P*-value**	**Fold change**	**MeV analysis**	**Multiple t-test per row *P*-value**	**Fold change**
hsa-miR-98	0.003	10.15	Up-regulated	0.137	8.4
hsa-let-7a*	0.010	9.33	Up-regulated	0.006	11.79
hsa-miR-194	0.019	8.12	No change	0.174	4.98
hsa-let-7f	0.018	6.63	Up-regulated	0.073	8.37
hsa-miR-22	0.043	6.40	Up-regulated	0.133	9.21
hsa-miR-15b	0.038	4.91	Up-regulated	0.001	5.37
hsa-miR-23b	0.030	4.87	Up-regulated	0.006	10.24
hsa-miR-21*	0.024	4.59	Up-regulated	0.059	6.02
hsa-miR-141	0.039	4.55	No change	0.507	3.16
hsa-miR-26a	0.020	3.98	Up-regulated	0.014	4.93

Table 3.5. Comparison of significant microRNA expression results with other team results. Comparison of significant miRNAs identified between cancer and control samples by *P*-value (Dunn's multiple comparison test) and fold-change in Dr D Guttery's group results (Palmieri et al., 2014) and the results found in this project (multiple t-test one per row with significant *P*-value <0.05 and fold change >2.5, and MeV software analysis). The two miRNAs* were investigated further in this study.

The selections of miRNAs were based on the results of pooled samples using microRNA Array analysing data. Eleven microRNAs (*miR-598, miR-30b, miR-652, miR-579, miR-671-3p, let-7a, miR-18a, miR-27a, miR-27b, miR-221, miR-21*) (Table 3.6) were selected for further study as potential cancer-specific miRNA biomarkers. These eleven miRNAs were selected as their expression levels in pooled cancer samples were significantly up-regulated compared to controls, according to our results and supported by the finding of Dr D. Guttery results.

Selected miRNAs	Comparison of miRNA expression levels in pooled cancer plasma samples (P1) and controls				
	Multiple t- test (our results)		MeV analysis	Dunn's test (Dr. D Guttery results)	
	***P*-value**	**Fold change**		***P*-value**	**Fold change**
hsa-miR-598	*0.0001**	3.66	No change	0.0001*	0.30
hsa-miR-30b	*0.0001**	4.02	No change	0.0002*	0.12
hsa-miR-652	*0.0005**	6.31	Up-regulated	0.00004*	0.38
hsa-miR-579	*0.0032**	9.52	Up-regulated	0.0001*	2.04
hsa-miR-671-3p	*0.0025**	11.21	Up-regulated	0.00004*	1.40
hsa-let-7a	*0.0058**	11.79	Up-regulated	0.010*	9.33
hsa-miR-18a	*0.0656*	6.85	Up-regulated	0.0183*	9.23
hsa-miR-27a	*0.0027**	7.89	Up-regulated	0.0002*	1.16
hsa-miR-27b	*0.0751*	9.18	Up-regulated	0.0121*	13.83
hsa-miR-221	*0.0376**	8.96	Up-regulated	0.0022*	6.09
hsa-miR-21	*0.0599*	6.02	Up-regulated	0.024*	4.59

Table 3.6. The selected miRNAs according to our results and Dr D. Guttery results. The selected miRNAs were significantly up-regulated in pooled cancer samples compared to controls, based on TaqMan microRNA array card results and according to 2 different analysis methods (multiple t-test one per row, fold change and MeV software analysis) and supported by Dr. D Guttrey results. These *miRNAs* were selected for further study in single Neocent plasma samples, as they could be potential cancer-specific miRNAs. *Significant *P*-value <0.05.

3.1.3 Comparison between baseline (P1) and end of treatment (P3) pooled plasma samples.

The previous TaqMan microRNA array data was also analysed to discover if there are miRNAs, which may serve as predictive biomarkers to predict response to the treatment.

Is there a difference in miRNA profile in pooled responder samples at baseline (RC-P1 and RE-P1) vs pooled samples after treatment (RC-P3 and RE-P3) in the same patients?

The miRNA expression profile in responder cancer samples at baseline (P1) and after treatment (P3) was analysed by two different methods [multiple t-test per row (*P<0.05*)

and MeV software] as before. According to the results, the expression of 13 *miRNAs* was significantly changed after treatment (P3) in compared to their expression at baseline (P1), seven miRNAs were up-regulated and six were down-regulated (Table 3.7). Three of down-regulated miRNAs were also found as down-regulated miRNAs in after treatment (P3) samples compared to (P1), with MeV software analysis (Figure 3.6; Table 3.7). None of these three miRNAs was significantly different in the baseline (P1) compared to controls except *miR-489* and *miR-429*, which was up-regulated, in baseline pooled plasma samples (P1) with multiple t-test one per-row.

MiRNAs	**Multiple t-test-one per row**				**MeV analysis**	
	P1 vs P3	***P*-value**	**Control vs P1**	***P*-value**	**P1 vs P3**	**Control vs P1**
hsa-miR-136*	↓	0.0011	-	-	↓	-
hsa-miR-597	↑	0.0026	-	-	-	-
hsa-miR-429	↑	0.0061	↑	0.0187	-	-
hsa-miR-219	↑	0.0084	-	-	-	-
hsa-miR-489*	↓	0.0107	↑	0.0095	↓	-
mmu-miR-153*	↓	0.0133	-	-	↓	-
hsa-miR-501	↑	0.0160	-	-	-	-
hsa-let-7f	↑	0.0242	-	-	-	-
hsa-miR-377	↑	0.0292	-	-	-	-
hsa-miR-376b	↓	0.0391	-	-	-	-
hsa-miR-502	↑	0.0392	-	-	-	-
hsa-miR-132	↓	0.0414	-	-	-	-
hsa-miR-106b	↓	0.0452	-	-	-	-

Table 3.7. Comparison of miRNA expression levels between responders before (P1) and after treatment (P3). The *miRNAs* with significant expression level change in responder samples after completing treatment (P3) compared to baseline pooled samples (P1). Also, the expression levels of miRNA in the baseline (P1) compared to controls. The comparison was based on ΔCt values, and analysed through Multiple t-test-one per row with a significant *P*-value <0.05 and according to MeV software analysis. The expression level change of miRNAs* was detected by two different analyses methods.

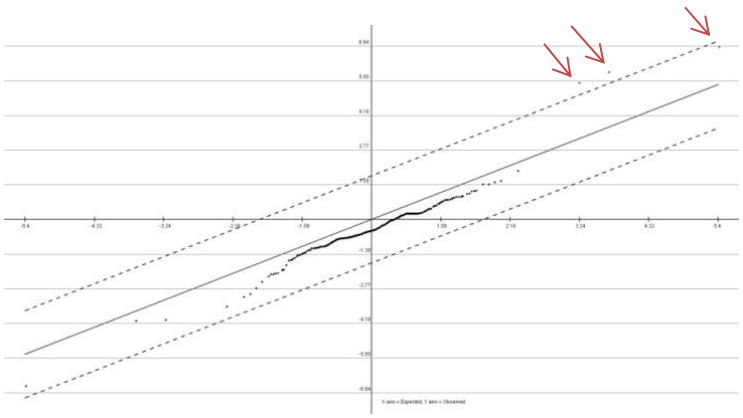


Figure 3.6. Comparison of miRNA expression levels between responders before (P1) and after treatment (P3), according to MeV results. The comparison of the expression level of 384 miRNAs between responder cancer samples at baseline (P1) and after completing treatment (P3), according to MeV software analysis. Three red dots which pointed by the arrows represent the down-regulated miRNAs. X axis =expected values and Y axis= observed values.

All miRNAs with significant expression change between responder samples at baseline (P1) and completing treatment (P3) were not selected for further study, their expression levels were very low in pooled cancer samples at baseline (P1) and after completing treatment (P3) (indicated by high Ct from TaqMan microRNA Array data), so the significant change in their expression levels could be inaccurate. Therefore, no further analysis was carried out on treatment response predictive biomarkers.

3.1.4 Validation of selected miRNAs in individual plasma samples

Having identified 11 candidates from TaqMan miRNA Array card data, real-time qPCR was processed on each cancer sample (P1) (n=35) and control samples (n=17) to validate the selected miRNA expression differences which were detected by the microRNA Array card in pooled samples. *MiR-16* and *miR-484* were used as reference genes in real-time qPCR as they were the most stable miRNAs by two different ways; manually, using TaqMan miRNA Array card data and by using GeNorm programme. Therefore, the average CT (cycle threshold) of *miR-16* and *miR-484* was used as reference genes in real-time qPCR to normalise any differences in CT of *miR-16* or *miR-484*.

The expression stability of *miR-16* and *miR-484* was validated through qPCR, in cancer samples (n=35) and control samples (Leicester controls, London controls and BSMS controls). The result showed no significant differences of *miR-16* or *miR-484* expression levels (based on CT mean) between control groups and cancer samples with *P*-value >0.05, except *miR-484* expression level which was significantly different in London controls compared to Neocent cancer samples with *P*-value=0.0002, (Figure 3.7) and that could be due to biological or technical differences. Also, the results showed a slight variation of *miR-16* or *miR-484* expression levels in different individual samples in the same control group or cancer which could be due to biological or technical differences (Figure 3.7).

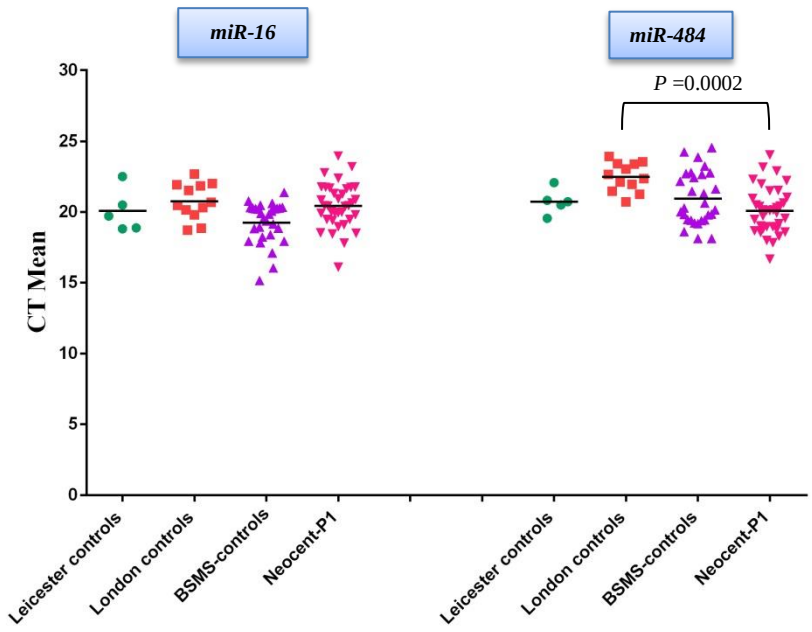
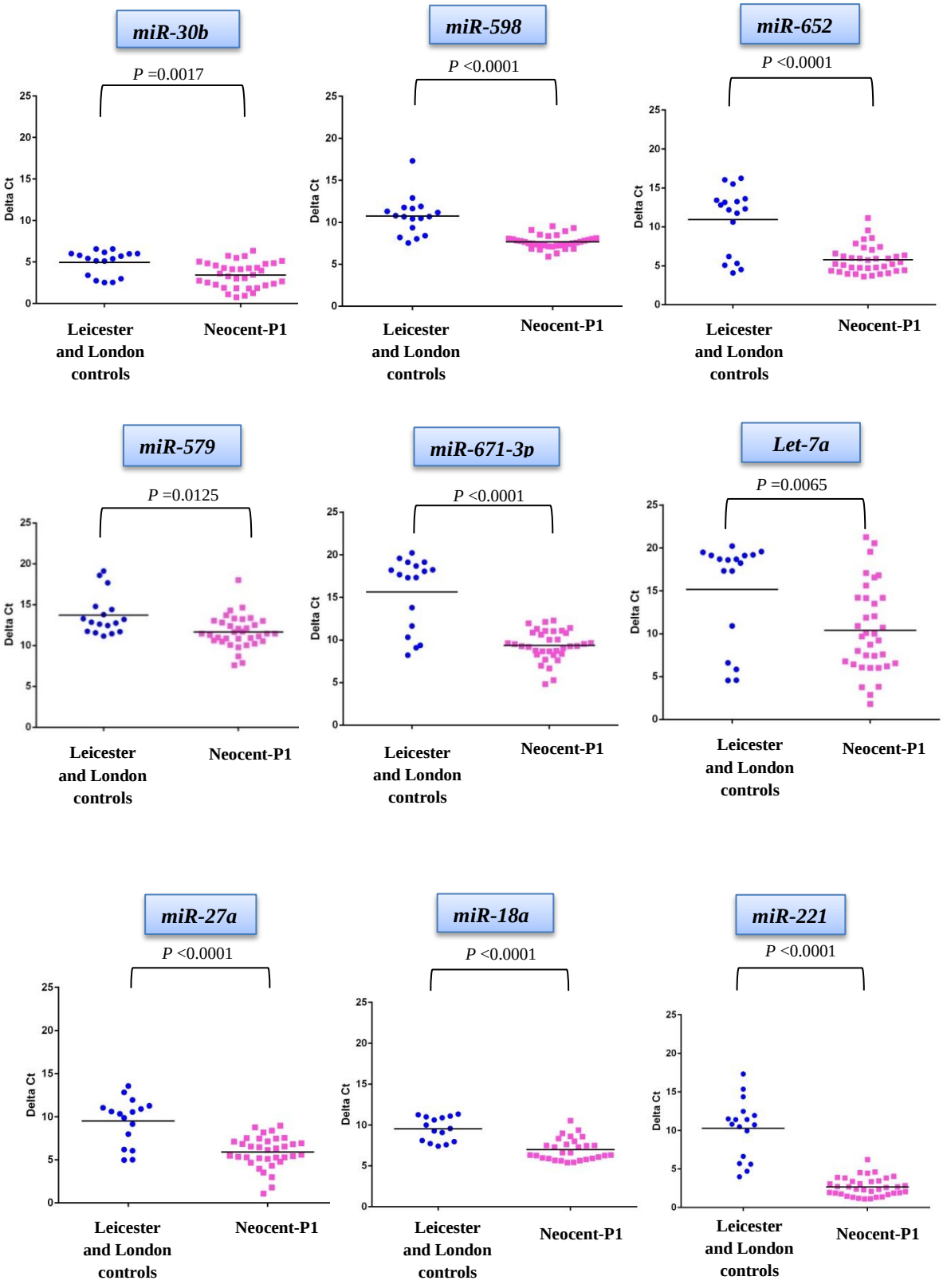


Figure 3.7. The *miR-16 and miR-484* expression level in different control and cancer groups. The expression profile of *miR-16 and miR-484* in single baseline Neocent cancer plasma samples (P1), n=35 and control samples (Leicester; n=5 and London controls; n=12 and BSMS; n=29) according to qPCR results. The horizontal line represents the mean of each group. The *P*-values represent the statistically analysis using one-way ANOVA (Tukey test) with a significance of $P<0.05$, to compare the *miR-16* or *miR-484* in different controls groups.

The qPCR was carried out on individual cancer plasma samples and individual control plasma samples (Leicester and London controls) to detect the expression level of the selected miRNAs. The results showed significant expression difference of all the selected miRNAs in individual cancer plasma samples compared to control samples, with $P<0.05$. All the selected miRNAs were up-regulated compared to control samples (Figure 3.8). There was an overlap of miRNA expression across cancer and control samples, but their expression differences were statistically significant. There was slight variation in the expression profile of all the selected miRNAs, except *let-7a* which showed wide variation in different Neocent cancer plasma samples, (Figure 3.8). Also, there was a wide variation in the expression profile of the single miRNA in different control samples (Leicester and London controls) (Figure 3.8). All of these findings indicate a wide variability in miRNA expression levels especially in control plasma samples as there were a group of outlier samples in most of the results (Figure 3.8). These expression level variations of single miRNAs in cancer plasma samples and control samples could be due to biological differences or processing differences,

although samples were collected and processed according to the same standard operating procedure (SOP).



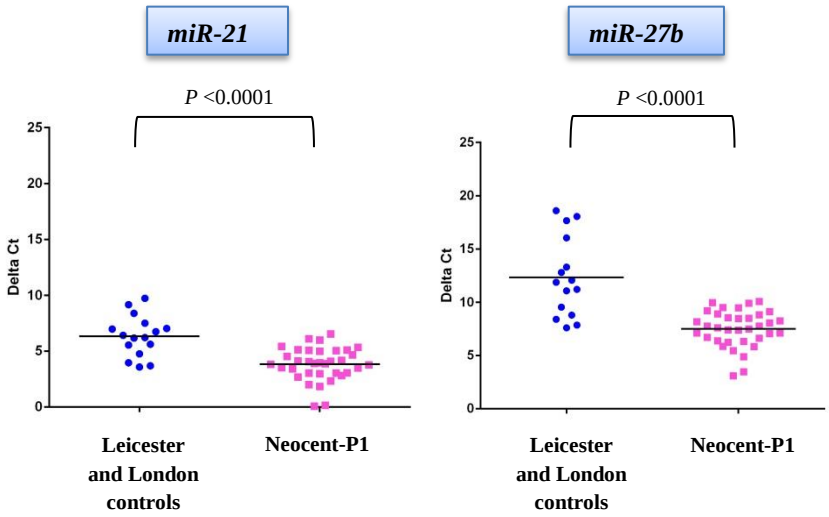


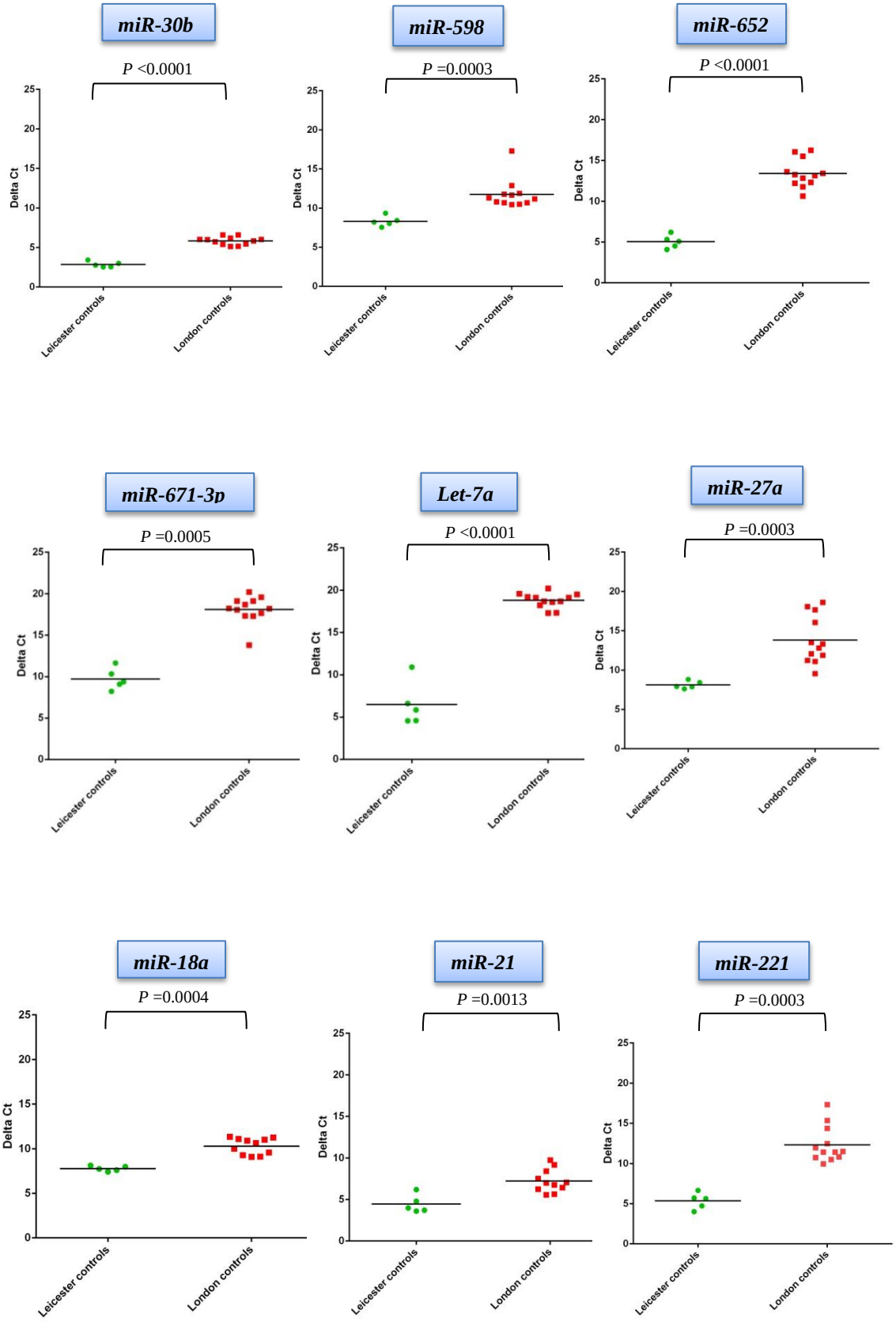
Figure 3.8. The selected miRNAs expression level in controls and cancer patients. The expression profile of selected miRNAs in single baseline Neocent cancer plasma samples (P1); n=35 and control samples (Leicester n=5 and London controls n=12), according to qPCR results. The horizontal line represents the mean of each group. The *P*-values represent the statistical analysis using un-paired t-test (Mann-Whitney test) with a significance of $P<0.05$.

The expression level difference of the same miRNA in two control groups (Leicester controls and London controls) was noticeable. Therefore, further validation was carried out to investigate if there was a significant difference in miRNA expression between the two control groups.

3.1.4.1 Validation of miRNA expression profile variation in different control plasma samples

The variation of specific miRNA expression level in control groups (Leicester and London controls) were validated through comparison of the selected miRNAs expression levels (*miR-598, miR-30b, miR-652, miR-579, miR-671-3p, let-7a, miR-18a, miR-27a, miR-27b, miR-221 and miR-21*) in two different control groups separately (Leicester controls and London controls) using qPCR and the *miR-16* and *miR-484* as reference genes. The result showed a significant difference ($P<0.05$) in the expression profile of almost all single miRNAs between two groups of controls (i.e. Leicester controls and London controls) with an exception of *miR-579* ($P<0.06$) (Figure 3.9). All the selected miRNAs which showed a significant difference in the expression profile between Leicester and London controls sample had a higher expression level (up-regulated) in Leicester control samples compared to London control samples (Figure 3.9). There was a wide variation in the expression level of some miRNAs in London

controls, such as *miR-27b* and *miR-579* (Figure 3.9), and that could be due to a small number of samples and/or biological differences.



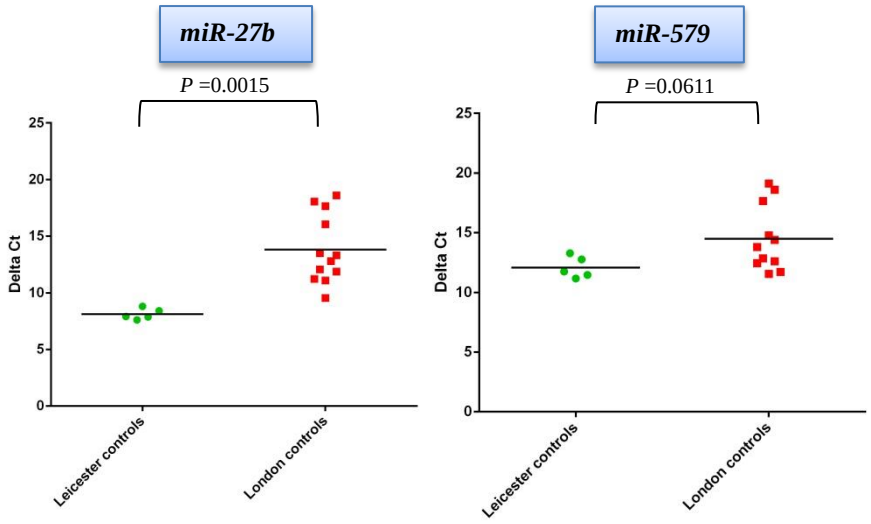


Figure 3.9. Selected miRNAs expression level in two control groups. The expression profile of the selected miRNAs in Leicester control and London control plasma samples according to qPCR results. The horizontal line represents the mean of each group. The *P*-values represent the statistically analysis using un-paired t-test (Mann-Whitney test) with a significance of $P<0.05$.

The expression difference of the same miRNA between two different control groups (Leicester controls and London controls) could be due to biological or processing differences. Although the samples were collected and processed according to the same standard operating procedure (SOP), but the only difference might be the timing between blood samples taken and processing the samples. There were 2 hours between collecting and processing Leicester controls samples, while London controls samples were collected during the clinic time, so they took up to 6 hours between collecting and processing the samples. These timing differences were investigated through:

(a) Each Leicester control sample was divided into two groups; one group processed after 2 hours and another group was left at room temperature for 6 hours before processing, to be similar to the London control group.

(b) New blood samples of 29 healthy female from BSMS study were processed in the same way as London and Leicester control groups. There were 2 hours between collecting and processing BSMS blood samples.

MiRNA was extracted from the newley prepared plasma samples (Leicester controls after 2 hours and 6 hours and BSMS controls) followed by qPCR of 8 miRNAs from the previously selected miRNAs (*miR-221, miR-652, miR-671-3p, miR-598, miR-27b, miR-21, miR-30b and miR-579*) with *miR16* and *miR-484* as an endogenous control.

The comparison of expression profiles of 8 miRNAs in 4 control groups (Leicester controls-2hours, Leicester controls-6 hours, London controls and BSMS-controls) showed a difference in London and 6 hours Leicester controls compared to 2 hours Leicester and BSMS controls (Figure 3.10). There was a slight variation in the expression profiles of a single miRNA in different samples in the same control groups.

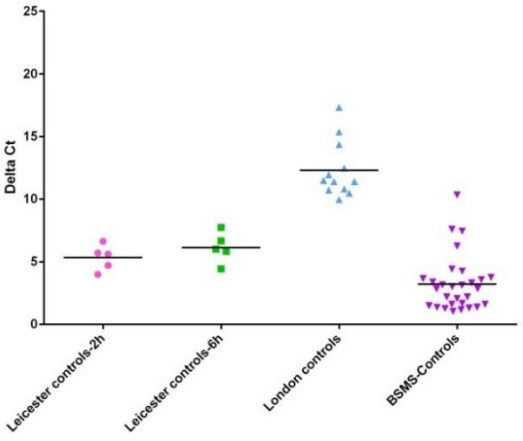
Is there any significant difference of the selected miRNAs expression profile between Leicester controls 6 hours and 2 hours plasma samples or Leicester controls 6 hours and London controls?

There was a significant difference with $P<0.05$, in the expression profile of 5 out of 8 miRNAs between Leicester control samples, left for 2 hours and 6 hours before processing (Figure 3.10; Table 3.8). These results may indicate that the timing between blood taking and processing could affect the expression profile of miRNAs in plasma. Although Leicester control-6 hours samples and London control samples left for 6 hours before processing, the expression profiles were significantly different in 5 out of 8 miRNAs (Figure 3.10; Table 3.8). These results may indicate that the timing between blood taking and processing could not be the only factor that affects the expression profile of miRNAs in plasma.

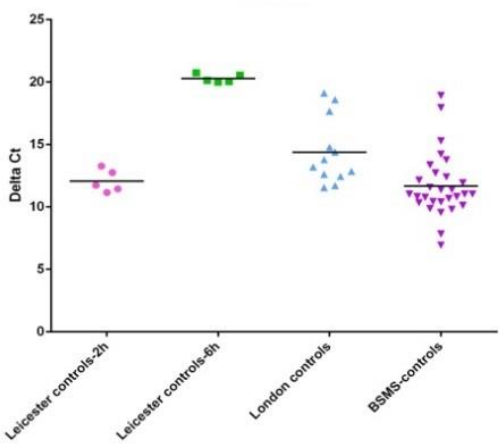
Is there any difference of selected miRNA expression profiles in BSMS control samples and other control groups?

The results of comparing the expression profile of the selected miRNAs between BSMS control samples and London control samples showed a significant difference in the expression profile of all the selected miRNAs with a *P*-value < 0.05 (Figure 3.10; Table 3.8). On the other hand, there was no significant difference in the expression profile of all the selected miRNAs in BSMS control samples and Leicester controls-2 hours, except *miR-221*, which showed a significant difference in its expression profile with $P=0.009$ (Figure 3.10; Table 3.8). The results of comparing the expression profile of the same miRNA between BSMS control samples and Leicester control-6 hours samples showed significant differences in most of the selected miRNAs (5 out of 8 miRNAs) (Figure 3.10; Table 3.8).

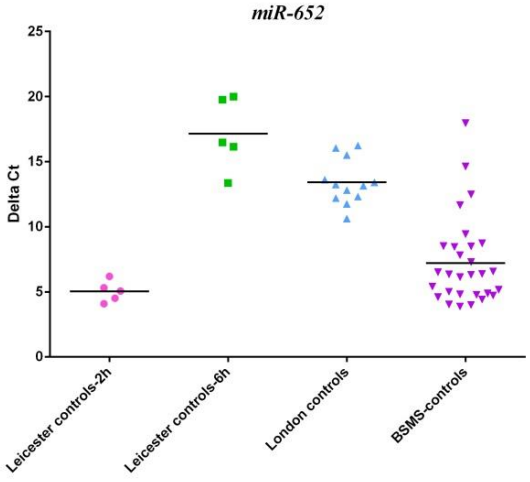
miR-221



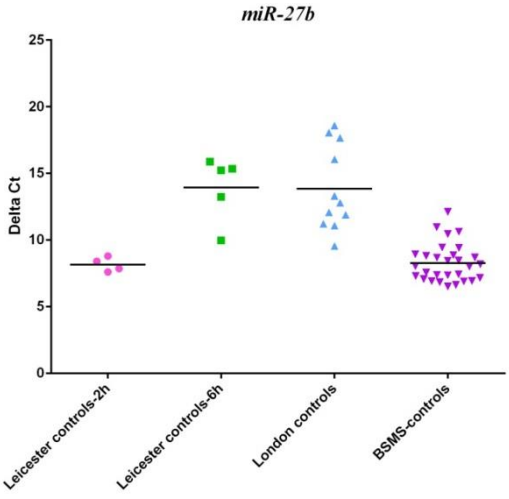
miR-579



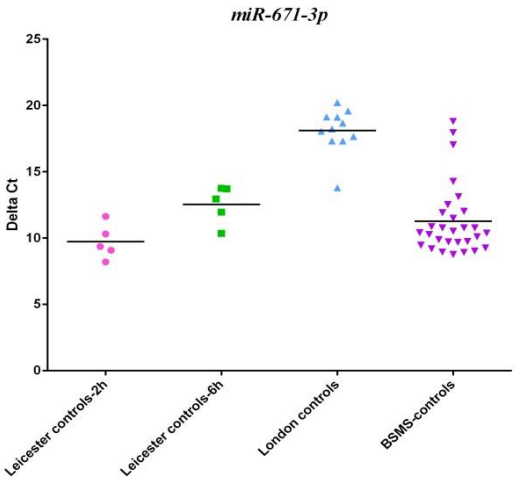
miR-652



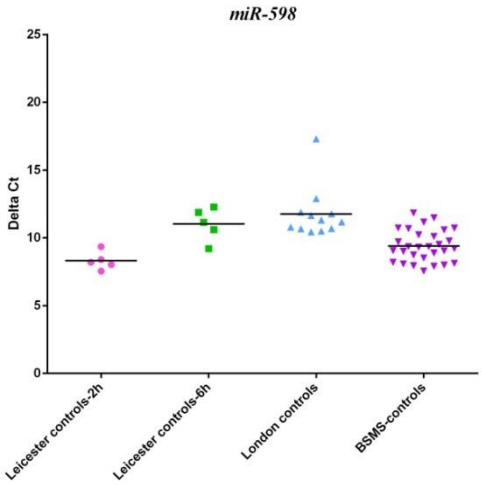
miR-27b



miR-671-3p



miR-598



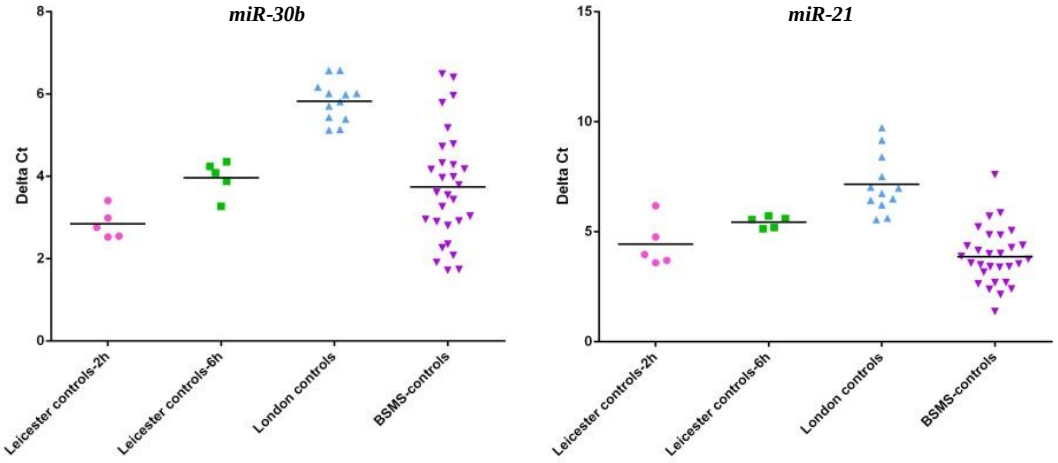


Figure 3.10. The expression levels of the selected miRNAs in different control groups. The expression profile of the selected microRNAs in 4 control groups [Leicester controls-2hours, Leicester controls-6 hours, London controls and BSMS-controls] according to qPCR results. The horizontal line represents the mean of each group.

Compared groups	Adjusted *P*-value							
	miR-221	***miR-652***	***miR-671-3p***	***miR-598***	***miR-27b***	***miR-30b***	***miR-21***	***miR-579***
Leicester controls-2h vs. Leicester controls-6h	0.222	<0.0001	0.0159	0.0129	0.0017	0.3684	0.5725	<0.0001
Leicester controls-2h vs. London controls	0.0003	<0.0001	0.0005	0.0003	0.0015	<0.0001	0.0013	0.0611
Leicester controls-2h vs. BSMS-controls	0.009	0.1266	0.1682	0.0591	0.8524	0.3292	0.7756	0.9854
Leicester controls-6h vs. London controls	0.0003	0.0094	0.0005	0.6745	0.9931	0.0117	0.0542	<0.0001
Leicester controls-6h vs. BSMS-controls	0.0037	<0.0001	0.0798	0.0187	<0.0001	0.9737	0.0526	<0.0001
London controls vs. BSMS-controls	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0085

Table 3.8. Comparison of selected miRNA expression levels between different control groups. The comparison of the selected miRNA expression profiles in control groups [Leicester controls after 2 hours, Leicester controls after 6 hours, London controls and BSMS-controls] using one-way ANOVA (Tukey test), with a significance of *P*-values >0.05.

In summary, the expression profiles of the selected miRNAs in BSMS control samples overlapped with Leicester controls-2 hours, but were significantly different with London controls and Leicester controls-6 hours and that could be due to processing differences as both London and Leicester-6 hours controls were left for 6 hours between collection and processing. However, there was a significant difference in expression level of some miRNAs between Leicester-6 hours and London controls, which may

indicate that the difference in miRNA expression level was not only affected by the time between collections and processing, but it could be other factors as the processing of the samples. As the expression levels of miRNAs were outliers in London control, the BSMS group was used as controls to compare with cancer samples, using qPCR.

3.1.4.2 Comparison of the selected miRNA expression levels between cancer and BSMS controls.

As there was a variation in miRNA expression levels between Leicester and London control groups and their small sample number, which may affect and bias the results when they are compared with cancer samples, the BSMS control group (n=29) was used to compare with Neocent cancer plasma samples. The qPCR was processed on each cancer plasma sample (35 Neocent plasma samples) and BSMS control samples (n=29) to validate the potential cancer-specific miRNA expression changes which were detected by the TaqMan microRNA Array in pooled samples. The expression levels of previously selected miRNAs (*miR-598, miR-30b, miR-652, miR-579, miR-671-3p, miR-27b, miR-221 and miR-21*) were examined through qPCR, as they showed significant differences in their expression level between cancer pooled samples and controls (Leicester and London controls) in TaqMan microRNA array data. The results showed no significant difference ($P>0.05$) in the expression profile of almost all the previous miRNAs (6 out of 8) between cancer plasma samples and BSMS controls (Figure 3.11). However, there was a significant expression change in *miR-598* and *miR-671-3p* with $P<0.0001$ and $P=0.004$ respectively (Figure 3.11). Both miRNAs (*miR-598 and miR-671-3p*) were up-regulated in cancer plasma samples compared to BSMS controls. However, their expression levels overlapped with BSMS control samples despite their statistically significant difference. Most of the miRNA expression levels in the BSMS control samples were slightly variable.

These findings were opposite to our previous findings, where all the selected miRNAs showed a significant difference in their expression between cancer and previous control groups (Leicester and London controls) and that could be due to wide variation of expression level of the selected miRNAs in Leicester and London controls, which could be due to processing and/or biological differences, and that raises the question that if the selected miRNAs were the right ones as they were selected based on the difference in their expression levels between cancer samples and control samples (Leicester and

London controls) from Taq Man microarray card results, despite some of selected miRNAs (e.g., *miR-21* and *miR-221*) have an oncogenic role and their potential use as biomarkers in breast cancer have been reported extensively (Chen et al., 2013; Singh & Mo, 2013).

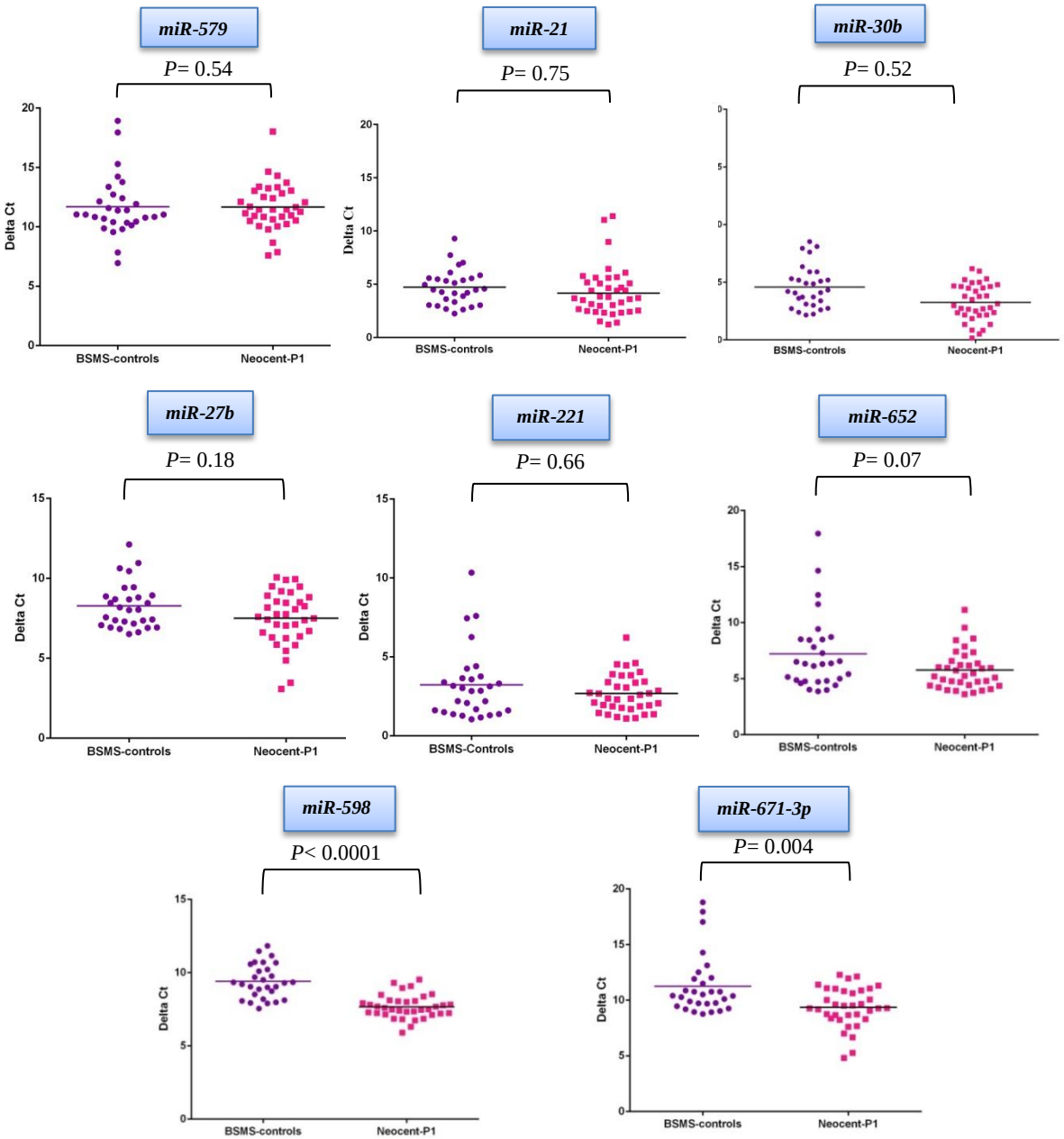


Figure 3.11. The expression levels of the selected miRNAs in the BSMS controls and cancer patients. The selected miRNA expression profiles in Neocent cancer plasma samples and BSMS controls, according to qPCR results. The horizontal line represents the mean of each group. The *P*-values represent the statistical analysis using un-paired t-test (Mann-Whitney test) with a significance of $P<0.05$.

3.1.4.3 Comparison between expression levels of miRNAs originating from plasma and exosomes

As there was wide variation in miRNA expression levels in control and cancer plasma samples, further validation of this variation was carried out through comparing of miRNA expression levels in plasma and exosomes, to investigate if miRNAs are more stable within exosomes and less affected by other processing factors which cause miRNA expression level variation in plasma. The expression level of the previously selected miRNAs (*miR-671-3p, miR-598, let-7a, miR-579, miR-652, miR-21, miR-30b and miR-27b*) was checked and compared between plasma and exosomes from Leicester control samples after 2 hours and after 6 hours. The qPCR was carried out on all samples, using *miR-16* and *miR-484* as reference genes. The result showed no significant differences in the expression levels of all miRNAs extracted from exosomes between Leicester control samples after 2 hours and after 6 hours (Figure 3.12) which may indicate that, the miRNAs extracted from exosomes are more stable as their expression level didn't affect by timing between sample collecting and processing. On the other hand, there were significant differences (P<0.05) in the expression levels of 6 out of 8 miRNAs (*miR-671-3p, miR-598, let-7a, miR-579, miR-652 and miR-27b*) extracted from plasma samples after 2 hours & 6 hours (Figure 3.12) as in our previous results, which showed that the variation of expression level of miRNAs extracted from plasma could be related to sample processing. These finding may indicate that miRNAs are more stable within exosomes than in plasma, especially with increasing the period of collecting blood to be left before processing them. However, the expression levels of some miRNAs were lower in exosomes compared to plasma, indicated by high delta Ct (e.g., *let-7a, miR-652 and miR-671-3p*) (Figure 3.12). These lower expression levels of miRNA in exosomes could be due to exosome depletion as a result of the isolation method of exosomes or may indicate a different origin of these miRNAs rather than exosome. Therefore, validation of the exosome isolation method is needed.

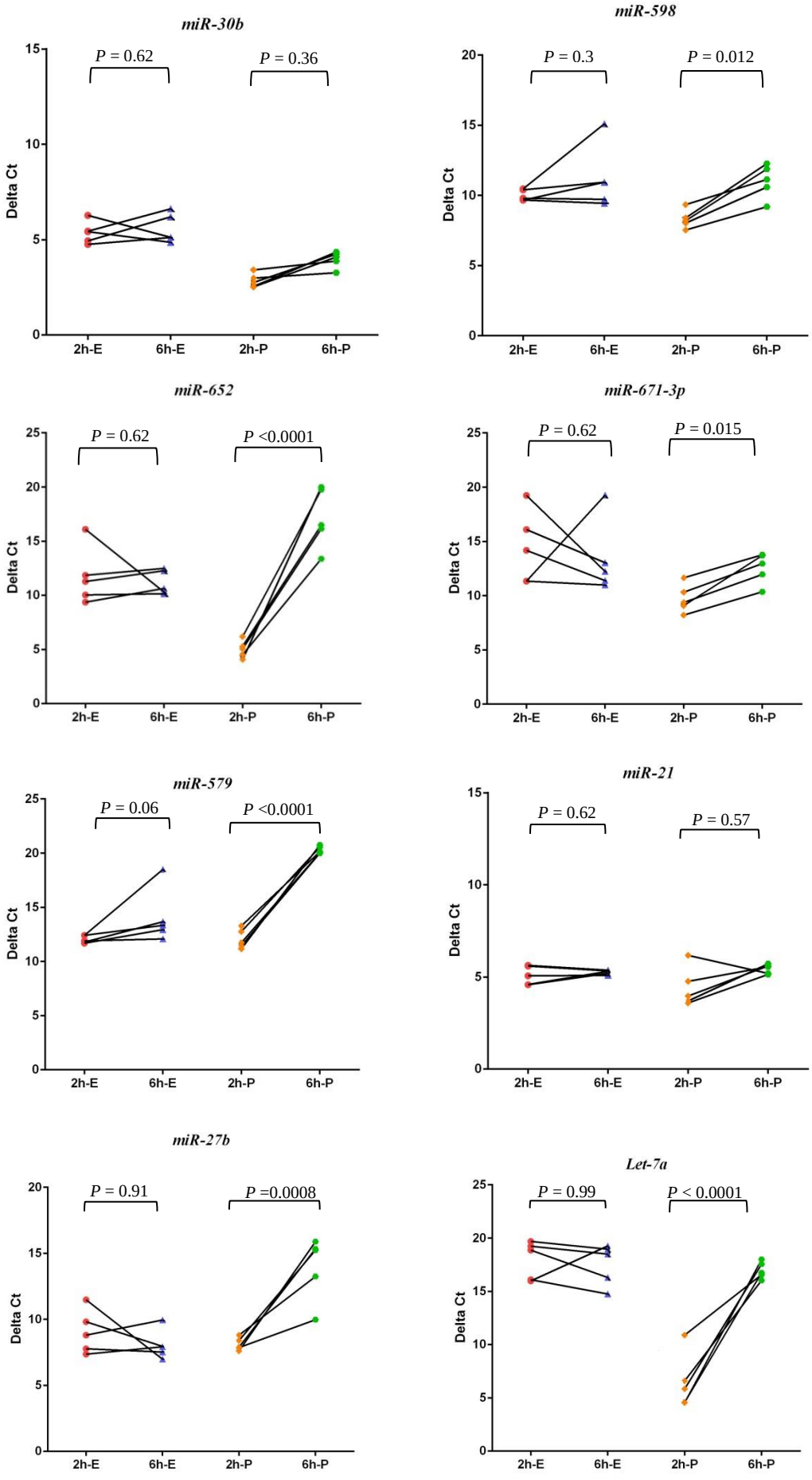

miR-30b
P = 0.62
P = 0.36
miR-598
P = 0.3
P = 0.012
miR-652
P = 0.62
P <0.0001
miR-671-3p
P = 0.62
P = 0.015
miR-579
P = 0.06
P <0.0001
miR-21
P = 0.62
P = 0.57
miR-27b
P = 0.91
P =0.0008
Let-7a
P = 0.99
P < 0.0001
Delta Ct
2h-E
6h-E
2h-P
6h-P

Figure 3.12. Comparison between expression levels of miRNAs originating from plasma and exosomes. Dots graph represent expression level of single miRNAs in Leicester control samples after 2 hours and after 6 hours (i.e., duration between blood taken and processing) which were extracted from exosome (E) and plasma samples (P). The line connected the delta Ct results of the matched samples which were processed after 2 hours and 6 hours. The *P*-values represent statistical comparison through un-paired t-test (Mann-Whitney test) in Leicester control after 2 hours or after 6 hours in samples extracted from exosomes or plasma, with a significant difference of $P < 0.05$.

3.1.5 The summary of miRNA expression results

Eleven miRNAs were selected as their expression levels in pooled cancer samples were significantly up-regulated compared to controls, based on the results of pooled samples using microRNA Array analysing data (Table 3.6). All the selected miRNAs were significantly up-regulated ($P<0.05$) in individual Neocent plasma samples compared to Leicester and London controls using qPCR (Figure 3.8). However, only *miR-598* and *miR-671-3p* out of the selected miRNAs were significantly up-regulated compared to BSMS controls using qPCR, with $P < 0.0001$ and $P=0.004$, respectively (Figure 3.11).

There was a significant difference in expression level of the same miRNAs between Leicester and London plasma controls (Figure 3.9), which could be due to time left between collections and processing of samples and/or other factors as the processing of the samples or biological factors. However, the expression levels of miRNAs extracted from exosomes were not effected by timing between collections and processing of samples (Figure 3.12). Therefore, the comparison between exosomes and plasma suggested that exosome derived miRNAs may be one way forward, but this was not investigated further, as these studies were performed towards the end of my project.

3.2 Discussion

Despite recent advances in cancer management and therapy, breast cancer remains the second leading cause of cancer death among women. Therefore, the identification of new stratification biomarkers is important to help with early diagnosis and management choices. But, the finding of a non-invasive approach for early detection of breast cancer is a challenging task. Identifying reliable blood biomarkers for early detection of different cancers is an area of intensive investigation. Proteins, DNA, and mRNA have been detected in the circulation of cancer patients (Anker et al., 1999; Petricoin et al., 2002) and several studies have reported the potential use of circulating biomarkers in

different body fluids, such as plasma and serum, as diagnostic and prognostic tools for different types of cancers. The miRNAs are a small noncoding RNAs that regulate gene expression at the post-transcriptional level (Bartel, 2004). Therefore, they play an important role in the regulation of most human genes, and they have been involved in different cellular processes including cell proliferation, survival, apoptosis, migration, invasion and differentiation (Bracken et al., 2015; Chang et al., 2015). During recent years, microRNA (miRNA) has become increasingly recognised as important regulator of both normal and cancer cell biology (He & Hannon et al., 2004; Hamam et al., 2014). Recently, miRNAs were emerged as potential biomarkers for disease status in a number of cancers as well as in other diseases, due to their stability and ease of detection (Chen et al., 2008; Mitchell et al., 2008; Creemers et al., 2012). For example, several studies demonstrated that some circulating miRNAs could discriminate breast cancer patients from healthy individuals (Cuk et al., 2013; Mar-Aguilar et al., 2013). Many miRNAs are up-regulated or down-regulated in tumours compared with normal tissues, supporting their different role in carcinogenesis as either tumour suppressors (e.g., *let-7a*) or tumour promoters called oncomirs (e.g., *miR-21* and *miR-221*).

Using TaqMan microRNA array cards, this study compared 384 different miRNAs expression levels in breast cancer plasma samples and age-matched healthy female plasma samples to discover potential cancer-specific miRNAs which could serve as novel biomarkers for early detection of breast cancer. The different analysis methods showed some different results: 69 out of 348 (18%) of circulating miRNAs expression level was significantly changed between cancer and controls with statistical analysis, while 87 out of 384 (22.7% or 23%) of circulating miRNAs were identified with significant expression change between cancer and controls with MeV software analysis. Overall, 49 out of 384 (12.8% or 13%) of miRNAs showed expression level changes between cancer plasma samples and controls with both analytic methods. All of these miRNAs were upregulated in cancer plasma samples compared to controls except one miRNA (*miR-509-5p*), which was down-regulated.

In Neocent breast cancer patients, *miR-509-5p* expression was reduced compared to the controls, supporting the finding of a previous study, which reported that *miR-509-5p* functioned as a tumour suppressor in breast cancer (Xing et al., 2015). The expression level of *miR-509-5p* was also down-regulated in other cancers including renal cell

carcinoma (Zhang et al., 2013) and non-small cell lung cancer (NSCLC) (Ma et al., 2016). Consistent with these findings, *miR-509-5p* has been suggested to play a major role in inhibiting cell proliferation and migration in renal cell carcinoma (Zhang et al., 2013) and cervical cancer and hepatocellular carcinoma (Ren et al., 2014). Also, *miR-509-5p* has been demonstrated to inhibit EMT-related genes in melanomas (Dong et al., 2015).

There was a high correlation between results from pooled and single cancer plasma sample, which was supported by qPCR results in the selected individual cancer plasma sample and controls. However, the expression level of the selected miRNAs showed no significant difference between cancer and the larger group of controls from the BSMS study, except *miR-598* and *miR-671-3p* which were upregulated in cancer patients. These different finding might be related to the broad variation of specific miRNA expression level in various control plasma samples, which could be due to processing and/or biological differences.

The vast variation in miRNA expression in different control samples may, in part, due to the effects of sample processing and other pre-analytic variables, such as miRNA profiles showed significant variation with increasing time before centrifugation (Page et al., 2013). Also, differences in sample processing and handling can be sources of considerable variation such as, plasma and serum processing (Duttagupta et al., 2011; McDonald et al., 2011), choice of anticoagulant (Kim et al., 2012) and hemolysis (McDonald et al., 2011; Pritchard et al., 2012) have been reported to affect miRNA measurement. Some studies suggesting that baseline blood counts impact circulating miRNA analysis (Duttagupta et al., 2011; Pritchard et al., 2012). As the majority of circulating miRNAs are expressed in one or more blood cell types was previously reported (Pritchard et al., 2012). Some studies provide specific practical recommendations to help minimise the variation attributable to plasma processing and platelet contamination as additional centrifugation (Cheng et al., 2013).

There was a significant difference in the expression profile of 10 out of 11 (90.9%) of the selected miRNAs between the Leicester and London controls. The timing between blood taking and processing could affect the expression profile of miRNAs in plasma, as miRNA profiles showed significant variation with increasing time before centrifugation (Page et al., 2013).

Although there was difference between Leicester and London controls which were used as controls for microarrays data analysis and miRNAs selection, most of the our results were supported by independent findings from Dr. D. Guttery. His research group carried out the same experiment independently (from plasma extraction through to TaqMan Array card) and found an overlap with these results (Palmieri et al., 2014), especially *let-7a* and *miR-21*. Also, most of the selected miRNA was also identified as an up-regulated or down-regulated in breast cancer compared to controls with other studies. For example, the higher expression level of *miR-21* was observed in a different type of cancer including breast cancer where the higher expression level of *miR-21* was found in the serum of patients with breast cancer or benign breast diseases compared to healthy controls (Schwarzenbach et al. 2012a). Also, *miR-21* was up-regulated in breast cancer plasma and tissue compared to normal (Ng et al., 2013). In our study, *miR-21* expression in Neocent breast cancer was up-regulated compared to Leicester and London controls, while there was no expression change compared to BSMS controls. These different results could be due to larger sample number in BSMS, biological or/and processing differences in control samples, although samples were collected and processed according to the same SOP.

The higher expression of *miR-21* was observed in an early and advanced stage of breast cancer with different studies as it was up-regulated in DCIS compared to normal (Chen et al., 2013; Li et al., 2014). Also, it showed higher expression level in primary breast cancer tissue especially in advanced tumour stage and lymph node metastasis and suggested as a prognostic biomarker for breast cancer (Yan et al., 2008).

The up-regulation of *miR-21* and *miR-221* in Neocent cancer patient compared to controls was supported by other studies as the oncogenic role of *miR-21* and *miR-221* and their potential use as biomarkers in breast cancer have been reported extensively (Chen et al., 2013; Singh & Mo, 2013). Moreover, other study observed the higher expression level of *miR-221* in plasma of hormone receptor negative breast cancer patients and in patients who respond to neoadjuvant chemotherapy (Zhao et al., 2011). While, *miR-221* was upregulated in the plasma of Neocent breast cancer patients which are Estrogen positive breast cancer. Therefore, the higher expression level of *miR-221* in breast cancer patient could not be related to the Estrogen receptor status.

The *let-7a* is well-known tumour suppressor miRNA that participates in the regulation of apoptosis, invasion and other cellular functions. The expression level of *let-7a* in Neocent cancer plasma sample was up-regulated compared to Leicester and London controls. However, there was no significant difference in the expression of *let-7a* between Neocent cancer plasma sample and BSMS controls, which are supported by the finding in a previous pilot study (Zhao et al., 2010). On the other hand, other study observed significantly elevated *let-7a* in breast cancer patients compared to age-matched and disease-free controls (Heneghan et al., 2010a; Palmieri et al., 2014). Conversely, other study found *let-7a* is down-regulated in breast cancer cells compared to normal breast cells (Zhao et al., 2011). These differences could be due to using of different study materials, as in our study and Zhao study used plasma for our analysis, while Heneghan study used the whole blood which contains different types of the cells, so the origin of miRNA could be circulating miRNA as well as cellular miRNAs. Also, this discrepancy could be due to heterogeneity of the breast cancer, as the studies may have different breast cancer stages. *let-7a* was significantly increased in patients who response to chemotherapy (Palmieri et al., 2014) and that could be due to inhibition of ERα-mediated growth in ER-positive breast cancer cells by *Let-7a,* which showed in previous studies (Yu et al., 2007; Sun et al., 2013). However, *Let-7a* expression doesn't significantly change in Neocent patients who response to chemotherapy.

The expression of *miR*-598 and *miR-671-3p* were the only two miRNAs, which were up-regulated in Neocent breast cancer tissue compared to both controls, which are BSMS controls and Leicester and London controls. However other study found different results, where *miR-671-3p* has been identified to be significantly down-regulated in other cancer such as, oesophageal adenocarcinoma versus normal and merely or not detectable in exosomes (Warnecke-Eberz et al., 2015). Also, the *miR-598* was down-regulated in breast cancer serum and tissue compared to normal (Zhu et al., 2014).

There was an up-regulation of circulating *miR-652* expression in Neocent breast cancer patient compared to London and Leicester controls which are in agreement with another study which showed increased levels of *miR-652* (Cuk et al., 2013). Conversely, other study indicated a decreased level of *miR-652* in patients with breast cancer compared to control (Mangolini et al., 2015), likely because these last two studies was performed on

serum instead of plasma samples. In support of this suggestion is the finding of Mangolini et al. when they repeated the experiment with miRNAs extracted from plasma rather than serum, they found an opposite trend of *miR-652* where it increased in breast cancer patients compared with controls (Mangolini et al., 2015). These finding of an opposite trend in serum or plasma samples is not new (Jarry et al., 2014; Ferracin et al., 2015) and raises the question about the different genesis of circulating miRNAs (Etheridge et al., 2013).

The expression of *miR-18a* was up-regulated in Neocent breast cancer tissue compared to Leicester and London controls and that was supported by another study which showed that, the expression of *miR-18a* was up-regulated in the serum of early diagnosed breast cancer compared to normal (Kodahl et al., 2014). Also, the expression of *miR-27a* and *miR-30b* were up-regulated in Neocent breast cancer plasma compared to Leicester and London controls and that supported by other study which observed up-regulation of *miR-27a* in breast cancer plasma and tissue compared to normal (Ng et al., 2013). However, other study showed down-regulation of *miR-27a* and *miR-30b* in breast cancer serum compared to normal (Luo et al., 2014).

In previous studies, specific circulating miRNAs, which were up-regulated in pre-treatment breast cancer plasma or tissue samples, they have been significantly reduced post-surgically, such as *let-7b*, *let-7g* and *miR-18b* (Cookson et al., 2012), *let-7a* (Palmieri et al., 2014) and *let-7a* and *miR-195* (Heneghan et al., 2010a). In Neocent plasma samples, the circulating *miR-136*, *miR-489* and *miR-153* were down-regulated after neoadjuvant treatment. However, there was no significant difference in the baseline (P1) expression levels of these miRNAs compared to controls, except *miR-489*, which were up-regulated. Also, their expression levels were very low in pooled cancer samples at baseline (P1) and after completing treatment (P3), so the change in their expression levels may not be significant.

The expression level of miRNAs extracted from plasma found to be significantly variable with increasing the time before the processing, which is supported by the finding of another study (Page et al., 2013). Therefore, miRNA was also isolated from exosomes. There was no significant variation in the expression level of individual miRNA extracted from exosome regardless of increasing the time before processing. However, the expression level of half of the exosome miRNAs was less than plasma

miRNAs, and that could be due to these miRNAs are less specific for exosomes and/or limited miRNA content in exosomes. Overall, this suggests exosomes may be more stable in terms of miRNA profiles, but this needs further investigation.

Chapter 4. *PIK3CA* and *TP53* somatic mutations analysis in breast cancer tissue and cf-DNA

4.1 DNA extraction for molecular analysis

Different methods of DNA extraction from breast cancer tissue were validated to extract good quality DNA which can be used for next generation sequencing (NGS) analysis. Seven FFPE slides containing the same metastatic breast cancer were used to compare extraction of the tissue DNA from FFPE tissue by two different methods: Phenol/Chloroform/IAA extraction and the GeneRead™ DNA FFPE kit (Qiagen kit). Three slides were used for each method, and one slide was stained with H&E (and reviewed by a consultant histopathologist) to mark the tumour tissue area.

The concentration of extracted DNA recovered using the two different methods was measured by qPCR using a *GAPDH* standard curve (Figure 4.1; Table 4.1).

	Quantity mean (10 ng/3.6 µL) (1:5 diluted DNA)	**DNA concentration (ng/µL)**
DNA-Phenol method	1.8	2.5
DNA- Qiagen Kit	1.76	2.45

Table 4.1. The concentration of metastatic breast cancer DNA. The concentration of metastatic breast cancer DNA extracted by the phenol method and the Qiagen kit using a standard curve.

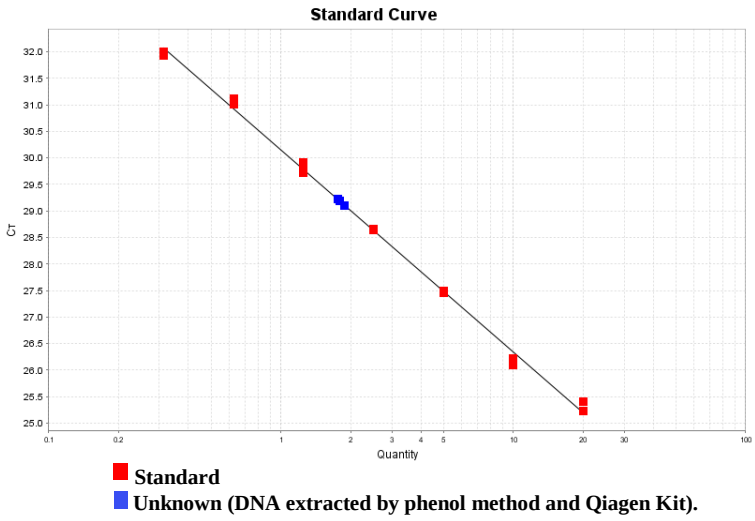


Figure 4.1. The concentration of metastatic breast cancer DNA extracted by the phenol method and Qiagen kit using a standard curve. The line graph shows the concentration of *GAPDH* in 1:2 HGD 7 serial dilutions (red colour) (standard curve) and metastatic breast cancer DNA extracted by the phenol method and the Qiagen kit (blue colour) through qPCR (R^2 =0.998, Effeciancy%=95.8%).

An equal amount of DNA was used for library preparation to prepare DNA samples for sequencing with Ion Torrent PGM. After library preparation, the amounts of DNAs were measured with a Qubit fluorometer. The concentration of library DNA extracted by the Qiagen kit was higher (1.34 ng/ µL) than the library DNA extracted by the phenol method (0.33 ng/ µL) and that could be due to better DNA quality when it extracted by the Qiagen kit.

The library prepared DNA was sequenced with the Ion Torrent PGM, and the results showed the reads from DNA extracted by Qiagen kit were much higher than from DNA extracted by Phenol method (Table 4.2). Again that could be due to better DNA quality when it was extracted by the Qiagen kit compared to the same DNA sample extracted by the phenol method. There were 12 variants detected in the sequencing results of DNA extracted by the Qiagen kit (Table 4.2). 5 out of these 12 variants were real somatic mutation using integrated genomic viewer (IGV) analysis (Table 4.3).

Sample	Mapped Reads	On target (compared with HG19)	Mean depth	Uniformity	Variants	Reads	Mean Read Length
DNA-Phenol	64,383	0.17%	0.44	93.63%	0	65,192	71 bp
DNA-Qiagen Kit	284,035	95.01%	1,073	57.65%	12	285,517	93 bp

Table 4.2. Sequencing results of DNA samples extracted by two different methods. Sequencing results of DNA sample obtained by two different methods (Phenol method and Qiagen kit) from the same metastatic breast cancer tissue. DNA was sequenced with the Ion Torrent PGM.

Chromosome	Position	Reference	variant	Frequency	Gene ID
chr7	55259559	G	A	3.7	*EGFR*
chr4	55561764	G	A	2	*KIT*
chr2	209113113	G	A	1.1	*IDH1*
chr4	153249384	C	T	1.1	*FBXW7*
chr17	7574003	G	A	1.1	*TP53*

Table 4.3. Somatic mutation in metastatic breast cancer DNA. Five somatic mutations in metastatic breast cancer DNA extracted by Qiagen kit, using Integrated genomic viewer (IGV) analysis.

4.2 *PIK3CA* and *TP53* Somatic mutation analysis

PIK3CA and *TP53* gene mutations are frequent in breast cancer (Stemke-Hale et al., 2008; Network TCGA, 2012). The top 3 hotspots of *PIK3CA* mutations in breast cancer

tissues are *PIK3CA* c.3140A>G, *PIK3CA* c.1633G>A and *PIK3CA* c.1624G>A respectively (www.mycancergenome.org) and the commonest three *TP53* somatic mutations are c.524G>A, c.818G>A and c.743G>A respectively (www.sanger.ac.uk) (Table 4.4). The presence of these common hotspots mutations in *PIK3CA* and *TP53* genes were chosen to examine in the cancer tissues DNA and their corresponding cf-DNA to find if they can potentially serve as screening biomarkers of breast cancer. However, detection of low-level somatic mutations in a high level of wild-type allele or cf-DNA requires the use of highly sensitive, specific and ideally quantitative methods. Several approaches were used to detect *PIK3CA* or *TP53* mutations for example qPCR, Sanger sequencing, HRMC, ddPCR and NGS. Validation of PCR, DNA Sanger sequencing and ddPCR in identifying low-level *PIK3CA* or *TP53* mutations was carried out to discover the very efficient and non-invasive tool in detecting low-level somatic mutations. The qPCR with PNA was performed for screening of *PIK3CA* mutations and HRMC method for screening of *TP53* mutations in the selected mutation hotspots. These methods were chosen as they are simple, quick and cheap. All the validations were carried out in cell lines DNA (Table 4.4), then applied to breast cancer tissues DNA and the avilable matched plasma DNA (Figure 4.2).

Cell lines	Gene name	CDS mutation	AA mutation	Exon	Mutation zygosity	Type
HCT-116 (Primary Carcinoma of colon)	*PIK3CA*	c.3140A>G	p.H1047R	20-exon	Heterozygous	Nonsynonymous substitution
MCF7 (Primary breast carcinoma, ER hormone+)	*PIK3CA*	c.1633G>A	p.E545K	9-exon	Heterozygous	Nonsynonymous substitution
SW948 (Primary adenocarcinoma of colon)	*PIK3CA*	c.1624G>A	p.E542K	9-exon	Homozygous	Nonsynonymous substitution
CCRF-CEM (Acute lymphoblastic leukaemia)	*TP53*	c.524G>A	p.R175H	5-exon	Heterozygous	Nonsynonymous Substitution
	TP53	c.743G>A	p.R248Q	7-exon	Heterozygous	Nonsynonymous Substitution
HT29 (Large intestine carcinoma)	*TP53*	c.818G>A	p.R273H	8-exon	Homozygous	Nonsynonymous Substitution

Table 4.4. Cell lines with different *PIK3CA* or *TP53* mutations. Cell lines with different *PIK3CA* or *TP53* mutations, their location, mutation zygosity status and mutation type.

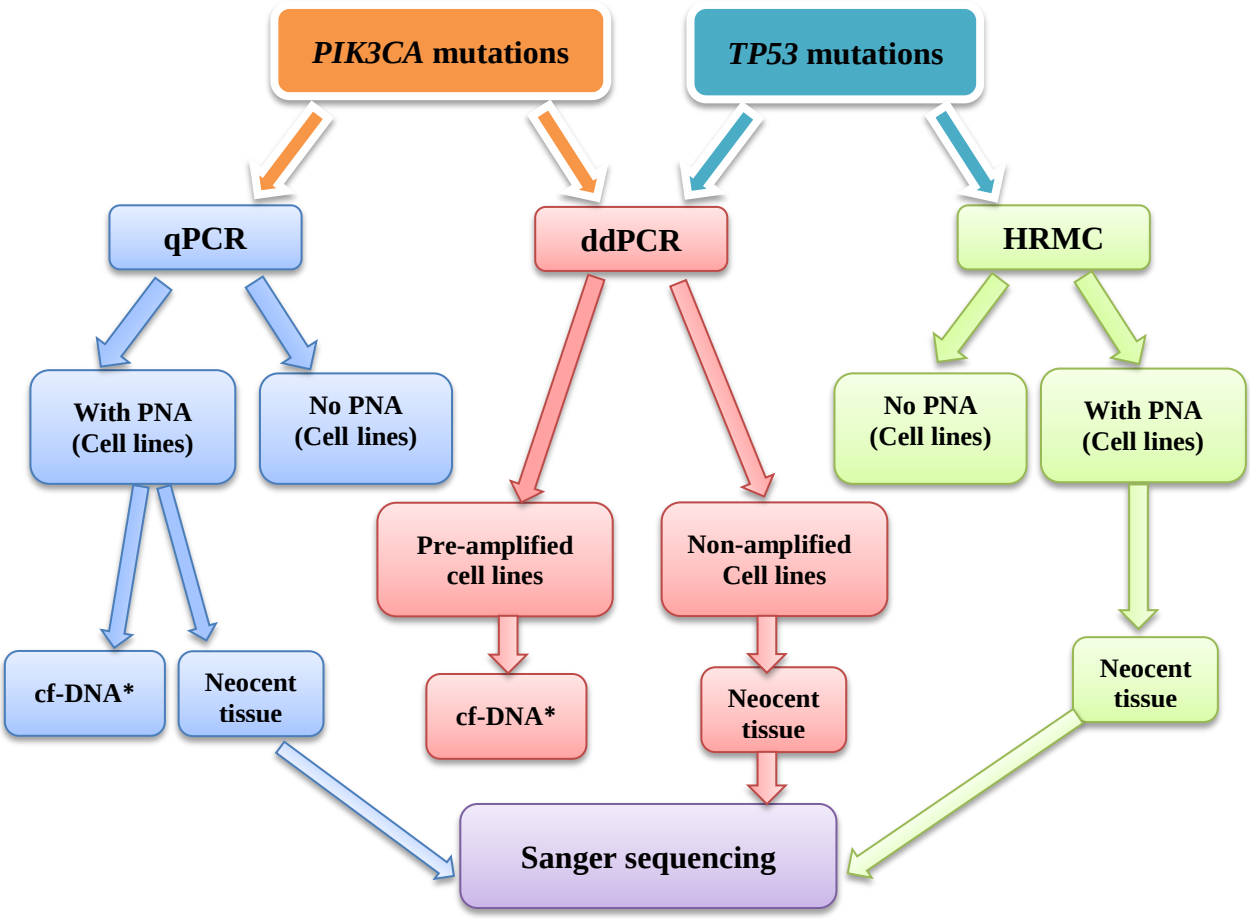


Figure 4.2. Workflow for *PIK3CA* and *TP53* mutations analysis using different methods.
* examination of cf-DNAs wasn't carry out due to the shortage of matched plasma sample available.

4.2.1 Comparison of various methods to detect *PIK3CA* and *TP53* somatic mutation in cell line DNAs

The *PIK3CA* and *TP53* mutations assays were designed using Primer 3 and Primer Express and validated to detect the low-level of *PIK3CA* or *TP53* mutations in cell lines DNA using ddPCR, qPCR with PNA, HRMC and DNA Sanger sequencing.

4.2.1.1 Digital droplet PCR (ddPCR)

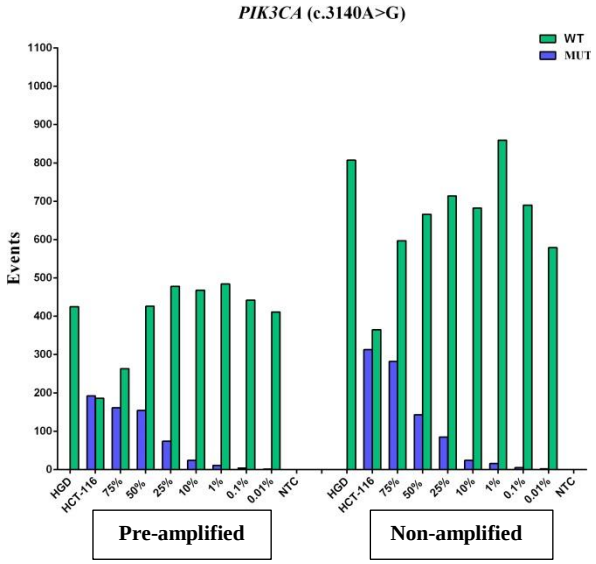
4.2.1.1.1 *PIK3CA* and *TP53* mutations detection in cell line DNA using ddPCR

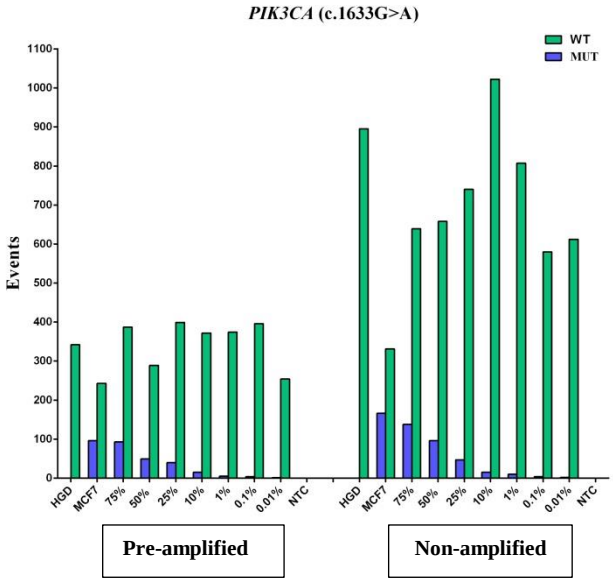
The *PIK3CA* and *TP53* mutations assays [forward primers, reverse primers and allele-specific probes (FAM/MGB Mutant probe and VIC/MGB Wild-type probe)] were

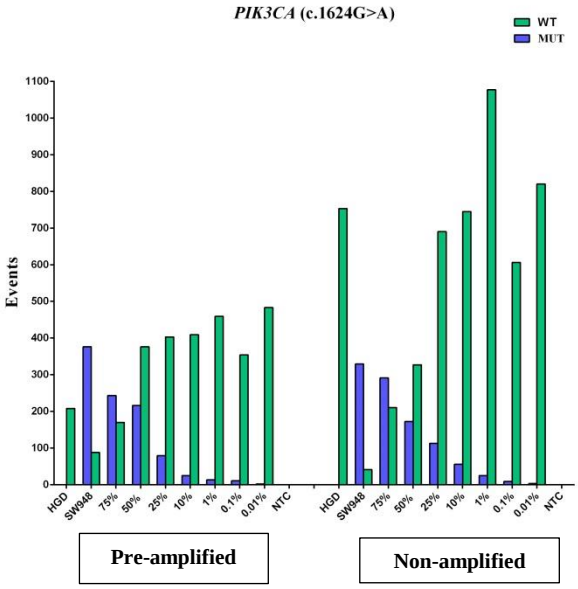
designed using Primer 3 and Primer Express and validated in cell line DNA, using the ddPCR method. The mutant allele was used as the target and wild-type allele as the reference. The six *PIK3CA* and *TP53* amplicons, which contain the selected hotspot mutation areas, were pre-amplified before using them as a template to validate ddPCR method in detection of *PIK3CA* or *TP53* mutations. The pre-amplification method was validated to be used with low DNA concentration as plasma DNAs (cf-DNA). The HGD, cell lines DNA and their dilution with HGD (75%, 50%, 25%, 10%, 1%. 0.1% and 0.01% of cell line DNA) were pre-amplified. Then the pre-amplified and the matched non-amplified DNAs were used as a templet to validate ddPCR method in detection of *PIK3CA* or *TP53* mutations. Also, compare the efficiency of mutant allele detection between pre-amplified and non-amplified DNA. HGD was used as a negative control, cell line DNA as a positive control and distilled water as non-template control (NTC).

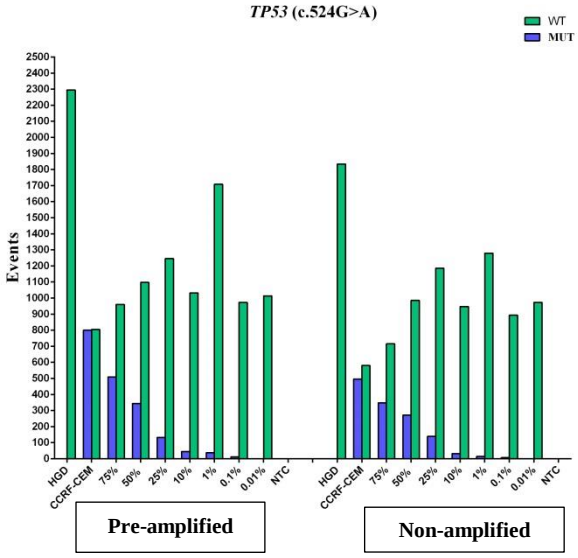
The ddPCR results of *PIK3CA* or *TP53* gene detection showed a variation in the number of droplets with wild-type or mutant alleles (Events) between pre-amplified and non-amplified DNA. These differences were more prominent with *PIK3CA* gene, as the number of droplets with *PIK3CA* wild-type or mutant alleles was higher in non-amplified DNA compared to pre-amplified DNA (Figure 4.3). The cut-off number of droplets to detect the mutation was ≥ 3 droplets with the mutant allele.

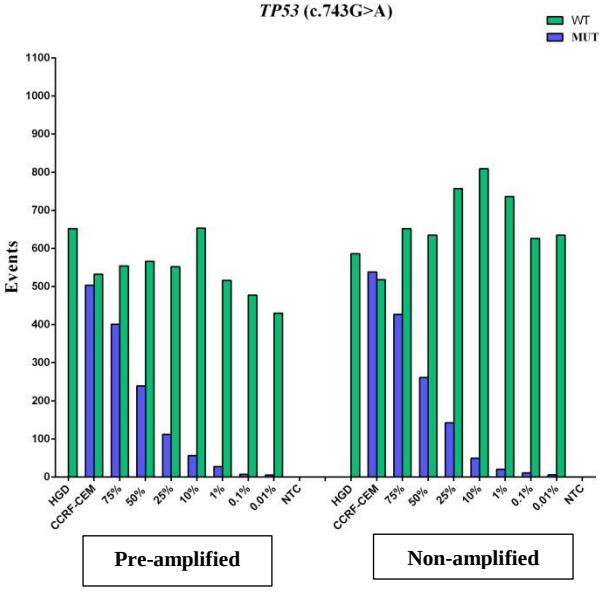
The ddPCR method was efficient in detecting the low concentration of the *PIK3CA* and *TP53* mutant alleles up to 0.1% of cell line DNA, in the pre-amplified and non-amplified diluted cell line DNAs, except *TP53* (c.743G>A and c. 818G>A) mutant alleles as they detected in up to 0.01% of cell line DNA (Figure 4.3).


PIK3CA (c.3140A>G)
WT
MUT
Events
HGD
HCT-116
75%
50%
25%
10%
1%
0.1%
0.01%
NTC
Pre-amplified
Non-amplified


PIK3CA (c.1633G>A)
WT
MUT
Events
HGD
MCF7
75%
50%
25%
10%
1%
0.1%
0.01%
NTC
Pre-amplified
Non-amplified


PIK3CA (c.1624G>A)
WT
MUT
Events
HGD
SW948
75%
50%
25%
10%
1%
0.1%
0.01%
NTC
Pre-amplified
Non-amplified


TP53 (c.524G>A)
WT
MUT
Events
HGD
CCRF-CEM
75%
50%
25%
10%
1%
0.1%
0.01%
NTC
Pre-amplified
Non-amplified


TP53 (c.743G>A)
WT
MUT
Events
HGD
CCRF-CEM
75%
50%
25%
10%
1%
0.1%
0.01%
NTC
Pre-amplified
Non-amplified

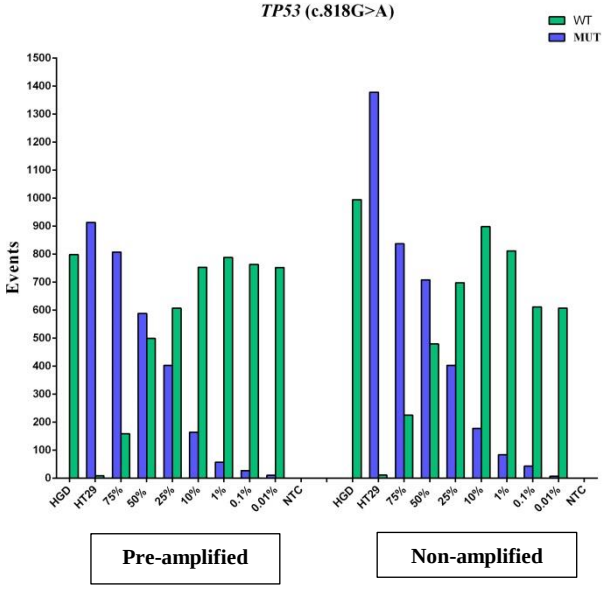
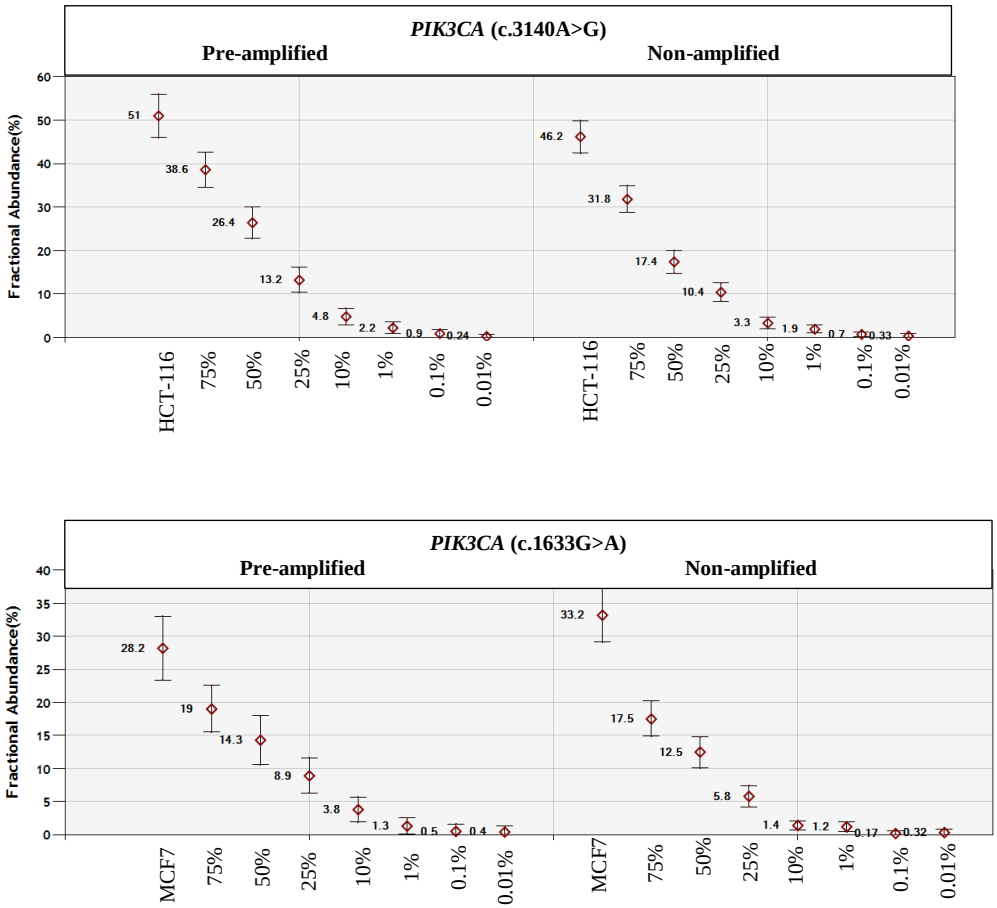
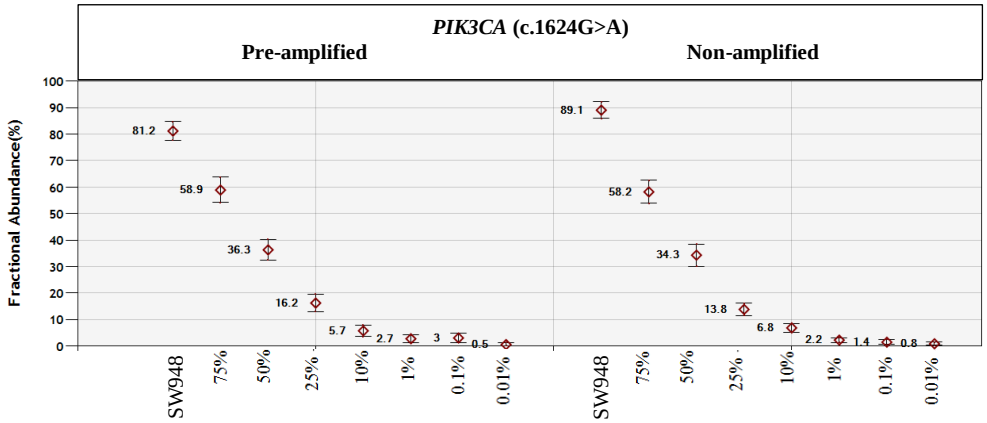
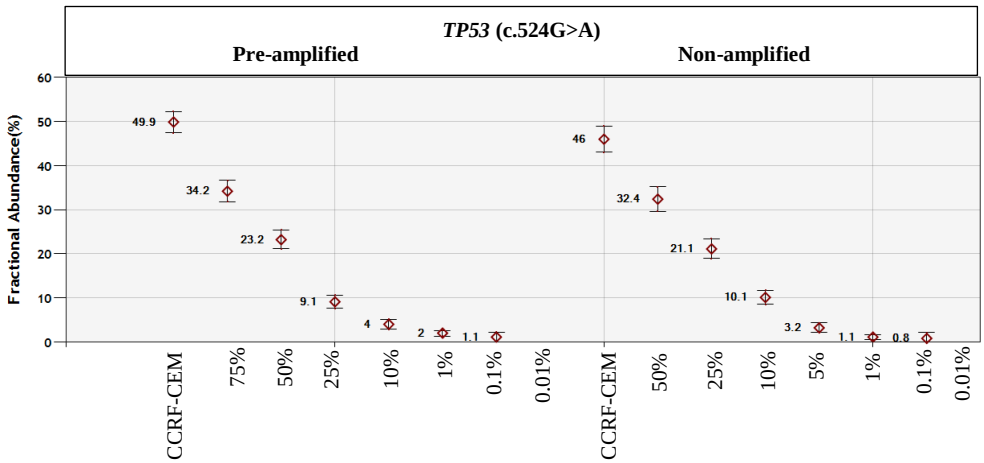

TP53 (c.818G>A)
WT
MUT
Events
HGD
HT29
75%
50%
25%
10%
1%
0.1%
0.01%
NTC
Pre-amplified
Non-amplified

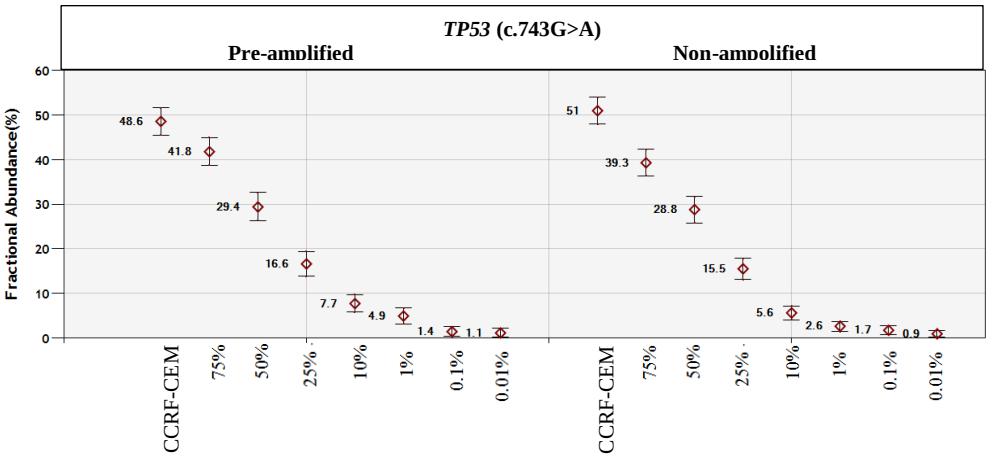
Figure 4.3. The comparison of *PIK3CA* or *TP53* mutation detection (Events) in pre-amplified and non-amplified cell line DNAs and HGD using ddPCR. Column graphs show the events which are the number of a positive droplet with mutant allele (blue column) and number of the droplet with wild-type allele (green column), in pre-amplified and non-amplified diluted cell line DNAs using ddPCR.

By comparing the fractional abundance (which is the percentage of droplets number with mutant allele to a total number of droplets with mutant and wild-type alleles), a lower fractional abundance was detected in non-amplified DNA compared to the paired pre-amplified DNA. The ddPCR method was efficient in identifying the mutation allele fraction ranging from 0.5% to 3% in pre-amplified DNA and 0.17% to 1.4% in non-amplified 0.1% of cell line DNA (Figure 4.4). While, *TP53* (c.743G>A and c. 818G>A) mutant alleles fraction was detected up to 1.1% to 1.5% in pre-amplified and 0.9% to 1.1 % in non-amplified 0.01% of cell line DNA (Figure 4.4).




PIK3CA (c.1624G>A)
Pre-amplified
Non-amplified
Fractional Abundance(%)
81.2
58.9
36.3
16.2
5.7
2.7
3
0.5
89.1
58.2
34.3
13.8
6.8
2.2
1.4
0.8
SW948
75%
50%
25%
10%
1%
0.1%
0.01%


TP53 (c.524G>A)
Pre-amplified
Non-amplified
Fractional Abundance(%)
49.9
34.2
23.2
9.1
4
2
1.1
46
32.4
21.1
10.1
3.2
1.1
0.8
CCRF-CEM
75%
50%
25%
10%
1%
0.1%
0.01%
5%


TP53 (c.743G>A)
Pre-amplified
Non-amplified
Fractional Abundance(%)
48.6
41.8
29.4
16.6
7.7
4.9
1.4
1.1
51
39.3
28.8
15.5
5.6
2.6
1.7
0.9
CCRF-CEM
75%
50%
25%
10%
1%
0.1%
0.01%

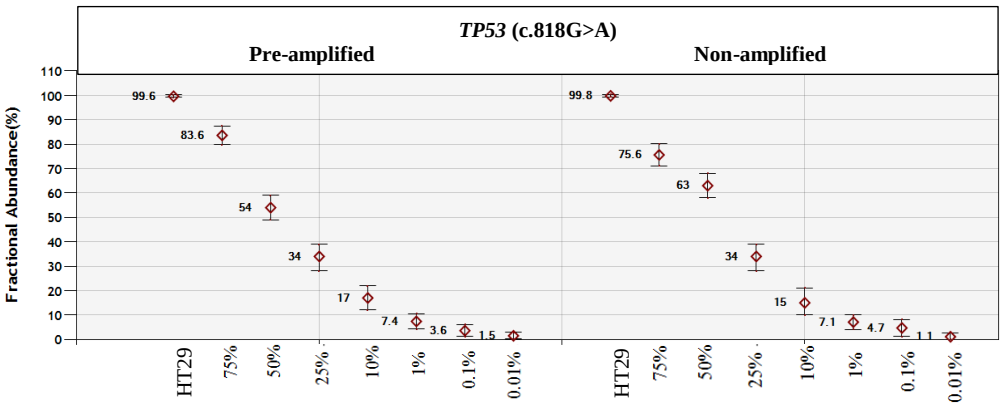
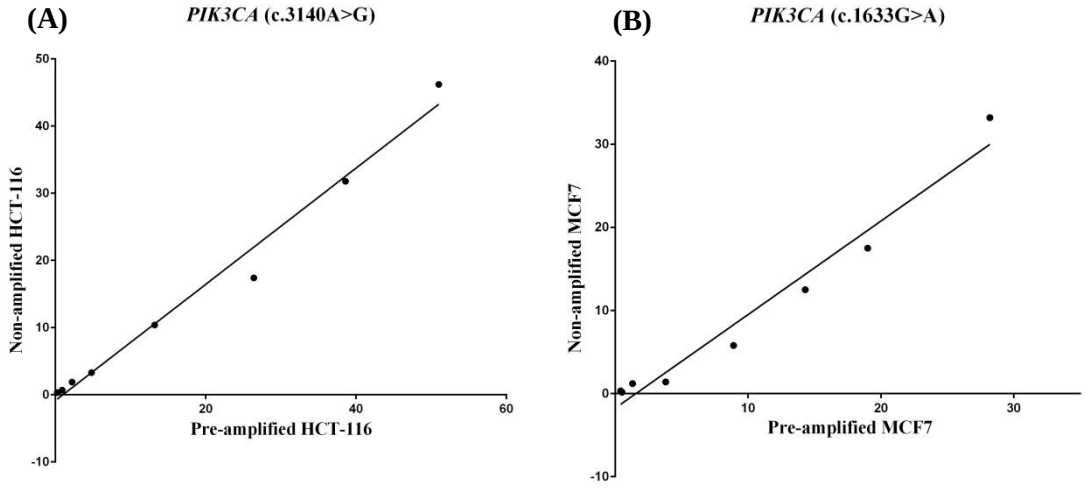

TP53 (c.818G>A)
Pre-amplified
Non-amplified
Fractional Abundance(%)
99.6
83.6
54
34
17
7.4
3.6
1.5
99.8
75.6
63
34
15
7.1
4.7
1.1
HT29
75%
50%
25%
10%
1%
0.1%
0.01%

Figure 4.4. The comparison of fractional abundance (%) of *PIK3CA* or *TP53* mutation in pre-amplified and non-amplified positive cell line DNAs. Dotted graphs show the percentage of droplets number with positive mutant allele to total number of droplets with and without mutant allele (Fractional abundance (%)), in pre-amplified and non-amplified diluted cell line DNAs with HGD (75%, 50%, 25%, 10%, 1%. 0.1% and 0.01% of cell line DNA), using the ddPCR.

There was no significant statistical difference between the fractional abundance (%) of the *PIK3CA* mutation in pre-amplified and non-amplified dilution of cell line DNAs, and there was a strong positive correlation between them, using Spearman test, (Figure 4.5; Table 4.5). Similarly, the fractional abundance (%) of the *TP53* mutation in pre-amplified and non-amplified dilution of cell line DNAs showed no significant difference (Figure 4.5; Table 4.5).

Somatic mutation		**Linear regression**		**Correlation**	
		R square	***P*-value**	**Spearman test (r)**	***P*-value (two-tailed)**
***PIK3CA* mutation**	*PIK3CA*-3140A>G	0.98	<0.0001	1.0	<0.0001
	PIK3CA-1633G>A	0.97	<0.0001	0.98	0.0004
	PIK3CA-1624G>A	0.99	<0.0001	0.98	0.0004
***TP53* mutation**	c.524G>A	0.997	<0.0001	1.0	0.0004
	c.743G>A	0.993	<0.0001	1.0	<0.0001
	c.818G>A	0.962	<0.0001	1.0	0.0004

Table 4.5. Correlation between fractional abundance (%) of non-amplified and pre-amplified DNAs. Correlation (Spearman test) and linear regression (R square) results between fractional abundance (%) values of non-amplified and pre-amplified diluted DNAs of HCT-116 (*PIK3CA*-c.3140A>G), MCF7 (*PIK3CA*-c.1633G>A), SW948 (*PIK3CA*-c.1624G>A), CCRF-CEM (*TP53* -c.524G>A and c.743G>A) and HT-29 (*TP53* -c.818G>A).



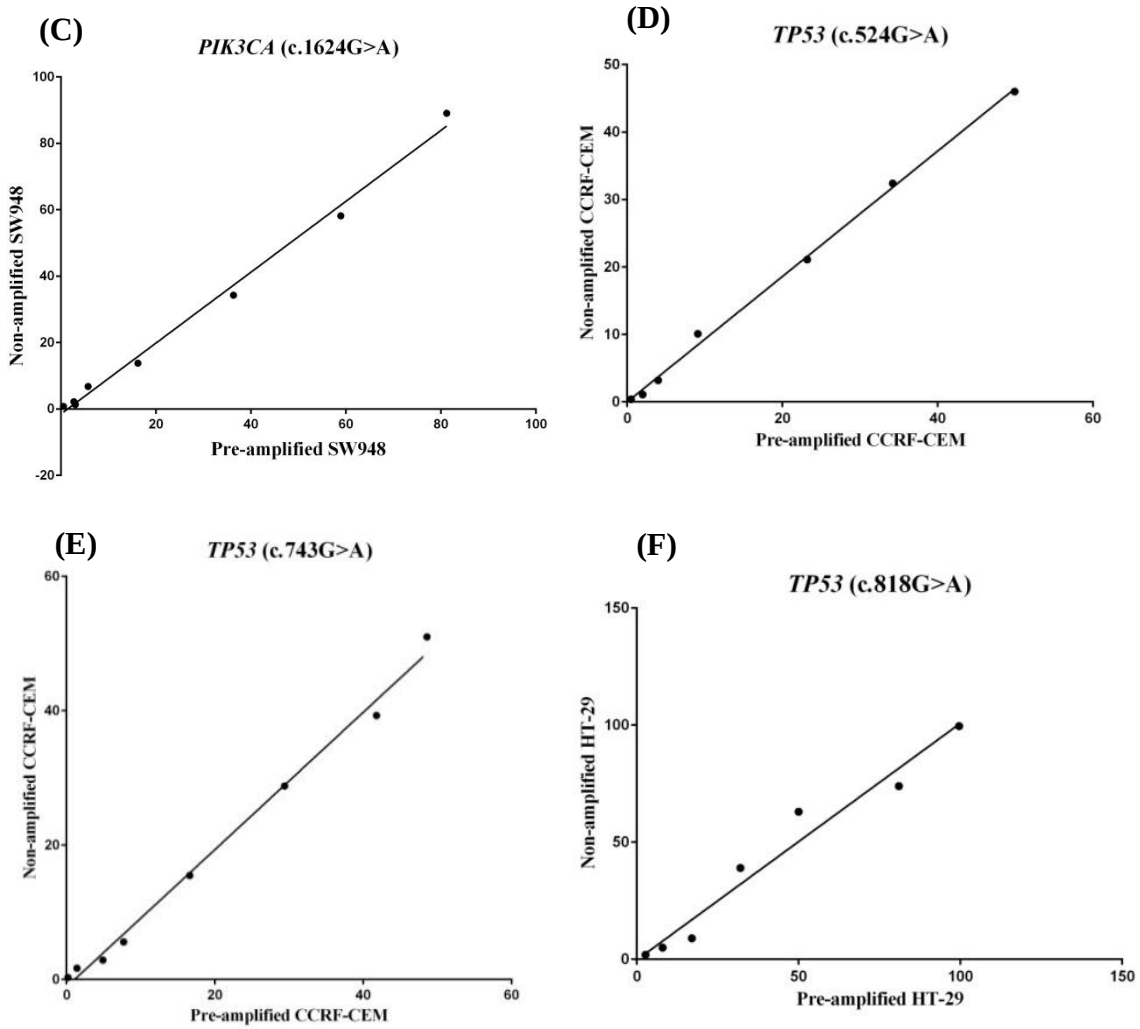


Figure 4.5. Correlation between fractional abundance (%) of non-amplified and pre-amplified diluted cell line DNAs. Dotted and line graph shows the correlation and linear regression between fractional abundance (%) values of non-amplified and pre-amplified diluted DNAs of HCT-116 with *PIK3CA* (c.3140A>G) mutation (A), MCF7 with *PIK3CA* (c.1633G>A) mutation (B), SW948 with *PIK3CA* (c.1624G>A) mutation (C), CCRF-CEM with *TP53* (c.524G>A) mutation (D), CCRF-CEM with *TP53* (c.743G>A) mutation (E) and HT-29 with *TP53* (c.818G>A) mutation (F).

4.2.1.1.2 *PIK3CA*-CNA (Copy number aberration) detection in cell line DNA using ddPCR

Changes in *PIK3CA* gene function could be related to the copy number aberration (CNA) within *PIK3CA* gene rather than mutation. Therefore, ddPCR method was validated to examine the presence of CNA within *PIK3CA* gene. The *RPPH1* gene was used as a reference gene to compare it with *PIK3CA* gene and examine the presence or absence of *PIK3CA* CNA. The forward primer, reverse primer and probe were designed in *PIK3CA* CNA region and *RPPH1* gene using Primer 3 and Primer Express programmes. The ddPCR method was validated to examine the presence of CNA within

PIK3CA gene in cell line DNA (RERF cell line) which contains *PIK3CA* CNA (www.sanger.ac.uk/COSMIC). HGD was used as negative control for *PIK3CA* CNA, RERF cell line DNA as positive control and its dilution with HGD (25%, 10%, 5%, 1% and 0.1% RERF DNA). The validation of ddPCR method, designed primers and probes in identifying the small concentration of *PIK3CA* CNA in higher concentration of HGD was carried out.

Based on HGD results, the normal ratio between *PIK3CA* gene copies to *RPPH1* gene copies was range from 0.8 to 1.2 (i.e., 1 +/- 0.2), and the normal diploid copy number of *PIK3CA* gene was range between 1.8 and 2.2 (i.e., 2 +/- 0.2). The ddPCR was efficient in detecting *PIK3CA* CNA in up to 10% of RERF cell line DNA, as its copy number was 2.51 and the ratio between *PIK3CA* gene copies to *RPPH1* gene copies was 1.26 (Figure 4.6A & B). On the other hand, the result of 5% RERF cell line DNA was borderline as its copy number was 2.27, but the ratio between its copies of *PIK3CA* gene and *RPPH1* gene was within the normal range and equal to 1.13 (Figure 4.6A & B).

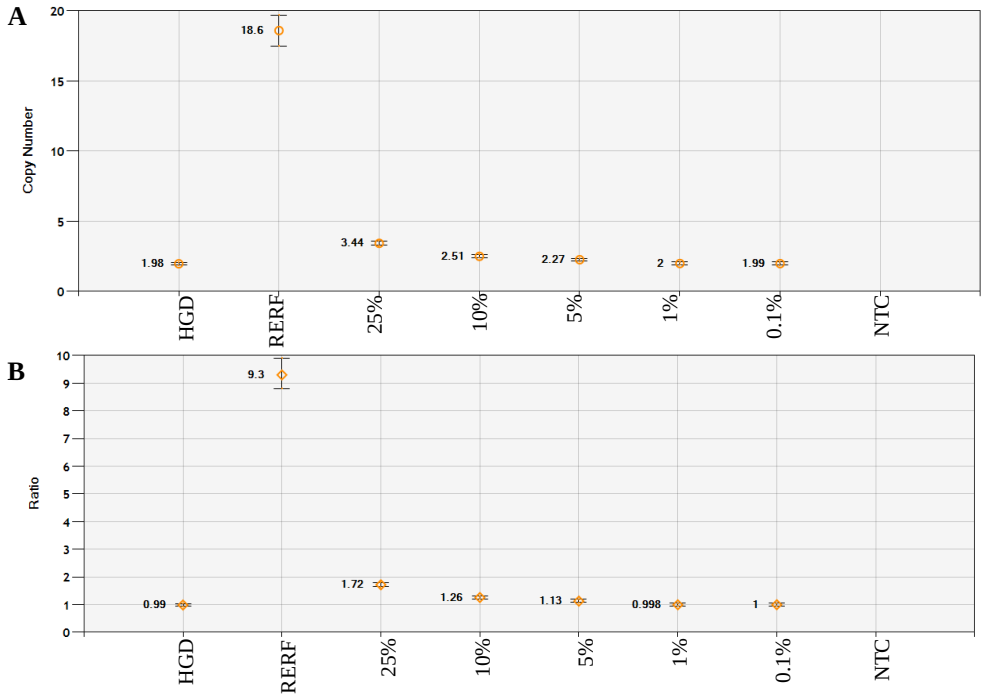


Figure 4.6. DdPCR results of *PIK3CA* copy number aberration in cell line DNA. *PIK3CA* Copy number (A) and the Ratio (B) between *PIK3CA* gene copies to *RPPH1* gene copies in HGD, RERF cell line DNA and its dilution with HGD (25%, 10%, 5%, 1% and 0.5% RERF DNA), using the ddPCR method.

4.2.1.2 The qPCR with PNA

Increasing the efficiency to identify low-level mutations could be achieved by blocking the wild-type alleles. Therefore, the sensitivity of peptide nucleic acids (PNA) in blocking the wild-type allele was assessed. The *PIK3CA* mutations assays [forward primers, reverse primers and allele-specific probes (FAM/MGB Mutant probe and VIC/MGB Wild-type probe)] were designed using Primer 3 and Primer Express. The PNA was designed using PNA BIO online programme. The sensitivity of the designed primers, specific allele probes (wild-type and mutant allele probes) and PNA were validated first in cell line DNA with *PIK3CA* mutations, and then in breast cancer tissue DNA and the available matched plasma DNA.

DNA of different cell lines with *PIK3CA* and their dilution with HGD (50%, 10%, 1%, 0.1% and 0.01% cell line DNA) were amplified by qPCR for 50 cycles, in duplex reaction (mutant and wild-type allele probes), with and without PNA, to validate the efficiency of PNA in blocking the wild-type allele and increasing the effectiveness of the mutant probe. HGD was used as a negative control and distilled water as a non-template control. The CT means of 45 or above was chosen to indicate no amplification.

The PNAs efficiently blocked the wild-type alleles in the regions of *PIK3CA* mutations (c.3140A>G, c.1633G>A and c.1624G>A), as there weren't any amplified amplicons with the wild-type allele (indicated by CT > 45) in HGD, cell line DNAs and their dilution with HGD (Figure 4.7A, B & C). Moreover, the presence of the PNA helped in detecting low level of mutant allele in 0.1% of cell line DNA compared to the reaction without the PNA as the mutant alleles (*PIK3CA* c.3140A>G and c.1633G>A) detected in up to 50% of cell line DNA and 10% of cell line DNA for *PIK3CA* (c.1624G>A) mutant allele without PNA (Figures 4.7A, B & C). Although, the presence of PNA helped in detecting low level mutation compared to the absence of PNA, the presence or absence of PNAs have insignificant effects in detection of the mutant allele in the presence of high level of mutation as in cell line DNAs and 50% cell line DNAs (Figures 7A, B & C). Moreover, the high level mutation (cell line and 50%) had slightly higher CT with PNA compared to the absence of PNA (Figure 4.7B & C) and that could be due to non-specific effect of PNAs where it may block a small concentration of mutant allele in the presence of a high concentration of mutant allele amplicons.

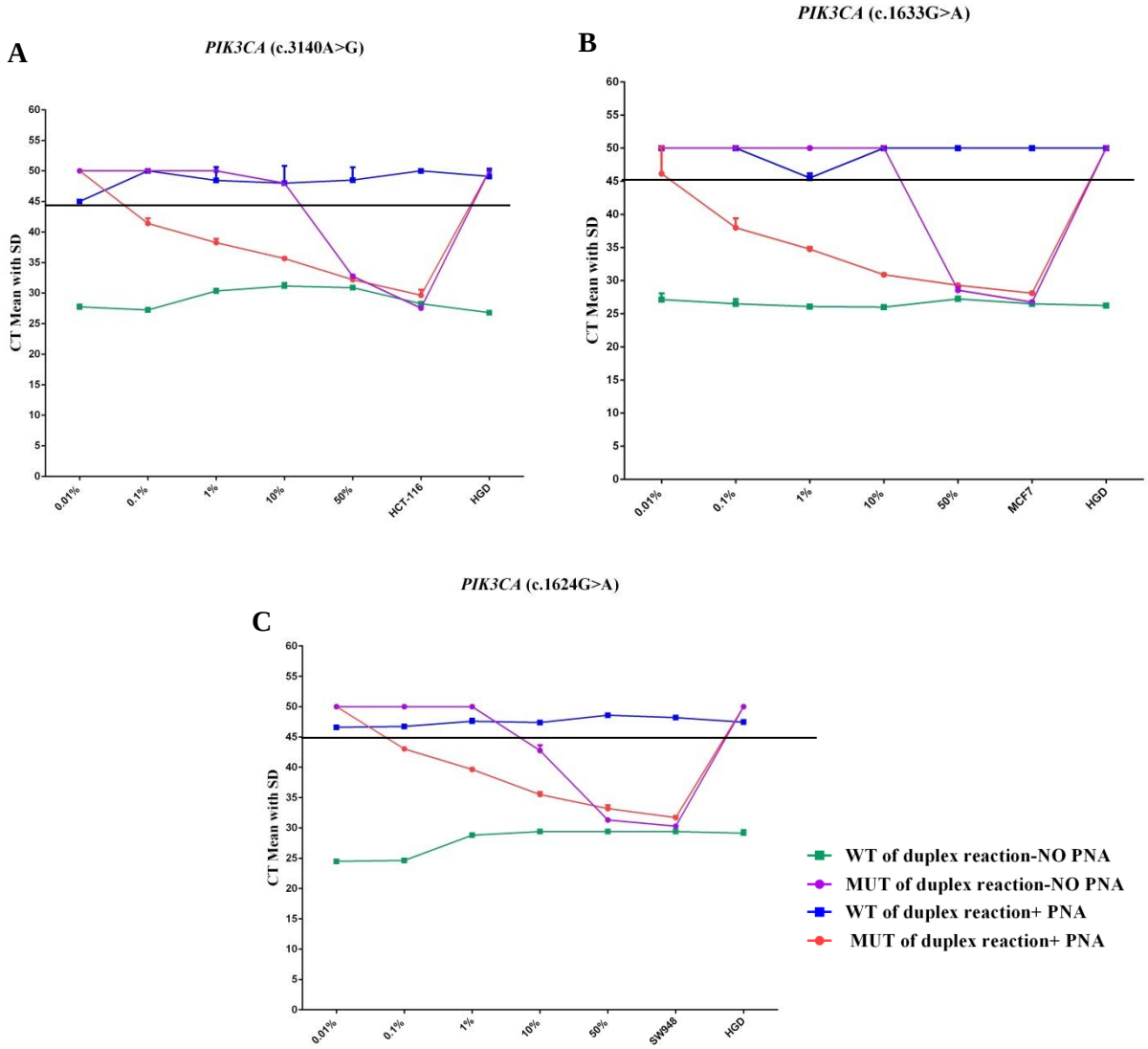
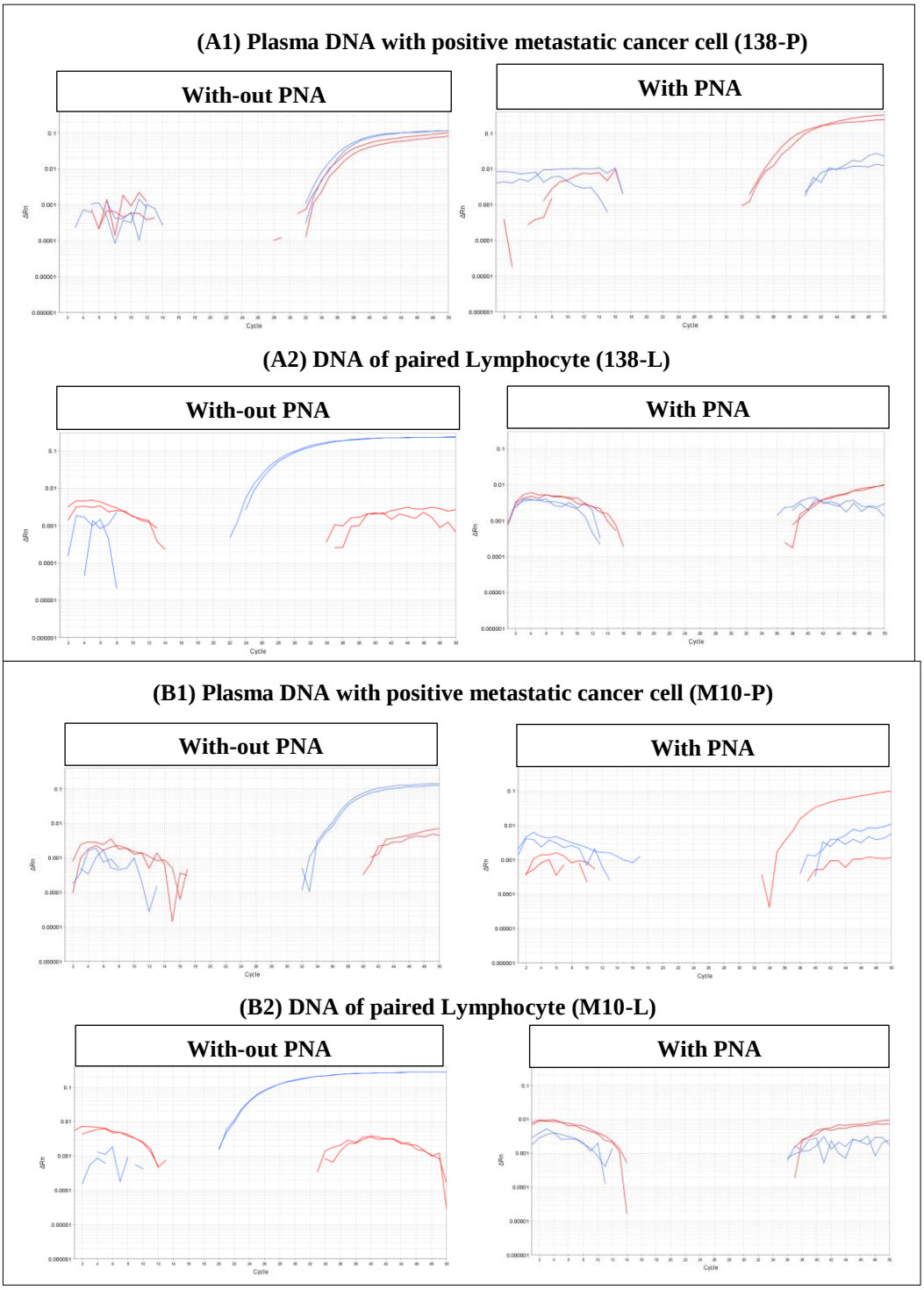


Figure 4.7. Comparison of *PIK3CA* mutations detection with and without PNA. The line and dotted graphs show CT mean with standard deviation (SD) of *PIK3CA* mutations (c.3140A>G, c.1633G>A and c.1624G>A) in HGD and dilution of HCT-116 (A), MCF7 (B) and SW948 (C) cell line DNA, through qPCR in duplex (i.e. mutant and wild-type probes) with and without PNA. The black line represents the CT mean amplification threshold= 45. MUT; represents mutant alleles and WT; represents wild-type alleles.

The validation of *PIK3CA* mutation primers, probes and PNA was also carried out in two plasma DNAs with *PIK3CA* (c.3140A>G) mutation and one plasma DNA with *PIK3CA* (c.1633G>A) mutation from metastatic breast cancer patients which detected by next generation sequencing (NGS). The presence of *PIK3CA* (c.3140A>G) mutation was detected in one out of two sequenced plasma samples, which were positive for this mutation, in the presence of PNA (Figure 4.8). While the presence of *PIK3CA* (c.1633G> A) mutation in sequenced plasma DNA couldn't be detected in the presence of PNA (Figure 4.9). The matched lymphocytes DNA were examined for *PIK3CA* mutation to confirm somatic mutation.

The second sequenced plasma DNAs with *PIK3CA* (c.3140A>G) detect *PIK3CA* (c.3140A>G) mutation in only one PCR reaction well from the duplicate reaction (Figure 4.8B1), and these results showed with repeated qPCR for the same DNA samples.



Wild type (*PIK3CA*-c.3140A>G)
Mutant (*PIK3CA*-c.3140A>G)

Figure 4.8. *PIK3CA* (c.3140A>G) mutation in sequenced metastatic breast cancer. *PIK3CA* (c.3140A>G) mutation examination in 2 sequenced plasma samples (138P and M10P) with *PIK3CA* (c.3140A>G) mutation (A1 & B1), using qPCR with and with-out PNAPNA. *PIK3CA* (c.3140A>G) mutation doesn't identify in their paired lymphocytes (A2 & B2). P; represents plasma DNA and L; represents lymphocyte DNA.

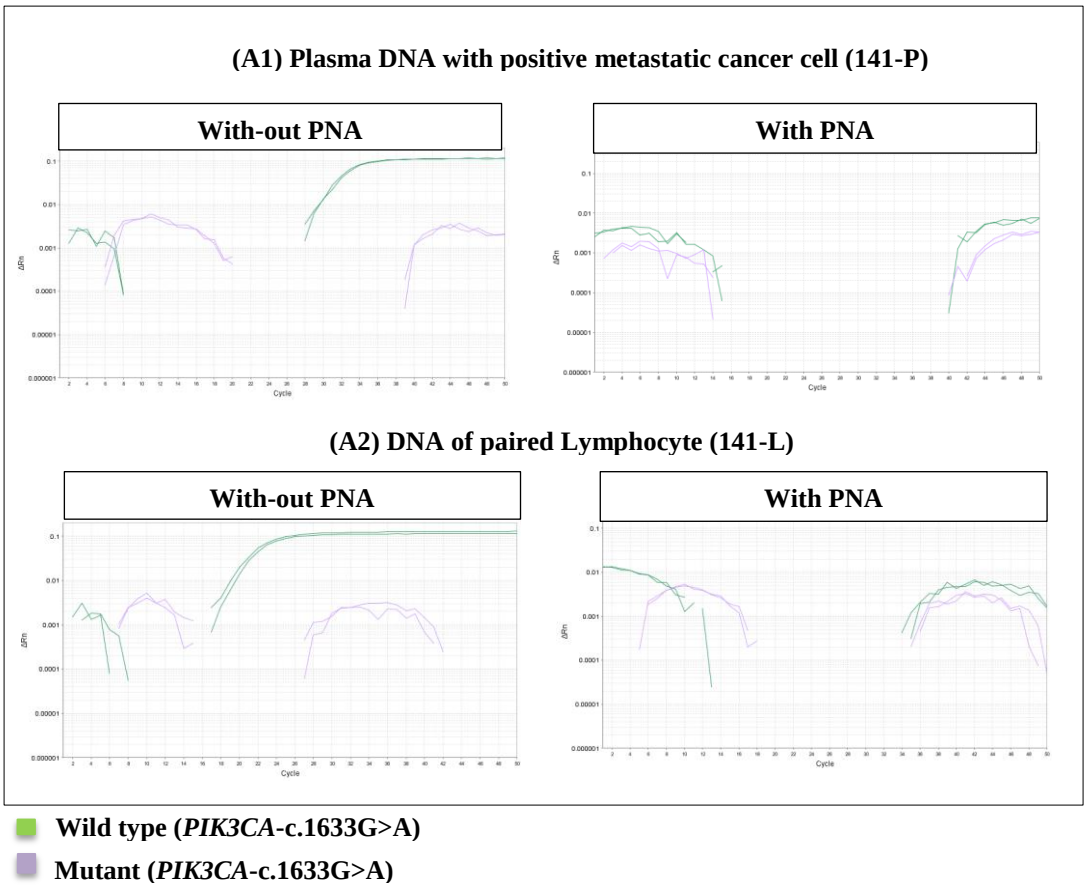


Figure 4.9. *PIK3CA* (c.1633G>A) mutation in sequenced metastatic breast cancer. *PIK3CA* (c.1633G>A) mutation examination in sequenced plasma sample with *PIK3CA* (c.1633G>A) mutation (A1) and its paired lymphocyte (A2 & B2), using qPCR with and with-out PNA.

After validation of *PIK3CA* mutation assays in positive cell line DNA and sequenced plasma DNA with *PIK3CA* mutation. The pilot study of 10 Neocent breast cancer tissues DNAs (N1T1, N2T1, N3T1, N6T1, N7T1, N8T1, N10T1, N11T1, N12T1 and N14T1) with whole genome amplification (WGA) were examined through qPCR, for 50 cycles, with duplex probes reaction and with PNA. The corresponding Neocent plasma DNAs with WGA at the baseline (P1) were also examined to validate the specificity of *PIK3CA* assay in detecting *PIK3CA* mutation in cf-DNAs. The matched lymphocytes DNA were tested for *PIK3CA* mutation to confirm somatic mutation.

The *PIK3CA* (c.3140A>G) mutation was detected in 4 out of 10 Neocent tissue DNA (N1T1, N3T1, N7T1 and N14T1). However, *PIK3CA* (c.3140A>G) mutation was not detected in their matched plasma DNA (cf-DNA) (P1) except in N3P1, which identified with higher CT (i.e., CT between 40 and 45) (Figure 4.10). The *PIK3CA* (c.1633G>A) mutation identified in one out of 10 Neocent tissue DNA (N2T1) but not in the corresponding plasma DNA (Figure 4.10). Lastly, *PIK3CA* (c.1624G>A) mutation wasn't detected in any of 10 Neocent tissue DNA samples. The matched lymphocyte DNAs of Neocent tissue DNAs, which were positive for *PIK3CA* mutations, were also examined, and all were negative for the *PIK3CA* mutations (Figure 4.10).

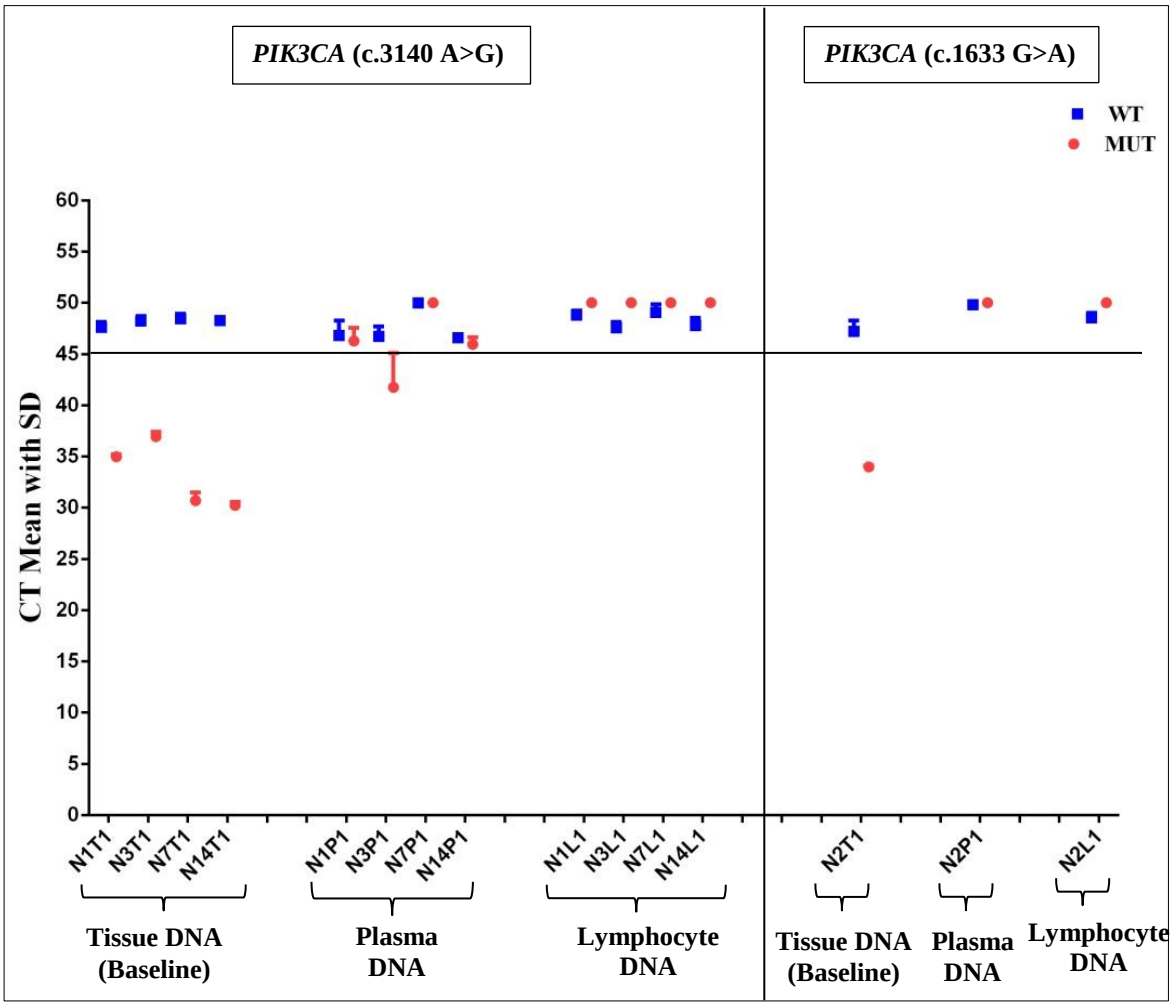


Figure 4.10. Detection of *PIK3CA* mutation in Neocent breast cancer tissues and their matched plasma and lymphocyte DNAs (Pilot study). Four out of 10 Neocent breast cancer tissues have *PIK3CA* (c.3140A>G) mutation and one have *PIK3CA* (c.1633G>A) mutation, indicated by CT < 45. Their matched plasma and lymphocyte DNAs were negative for *PIK3CA* mutation, except for N7P1. Black line represents the CT mean amplification threshold = 45 (CT mean < 45 indicates the amplification).. The blue squares represent wild-type alleles (WT), and the red circles represent mutant alleles (MUT).

4.2.1.3 High resolution melting curve method (HRMC)

The HRMC method was chosen for detection of *TP53* mutations as the mutations are distributed in most of the coding exons of the *TP53* gene and they composed of GC-rich regions (www.cancer.sanger.ac.uk/COSMIC), which make specific assays design more challenging. Also, the HRMC method reveals the difference in the melting temperature between DNA with mutant allele and wild-type DNA (Reference DNA). Therefore, HRMC method is not specific for a particular mutation but can detect any mutations within the amplicon. The PNA approach was also used with HRMC as its efficiency in blocking the wild-type allele could lead to increase the effectiveness of detecting the low-level mutation.

The *TP53* mutations assays [forward primers, reverse primers and general probe (FAM/MGB)] were designed using Primer 3 and Primer Express and validated in cell line DNAs using HRMC with PNA for 40 cycles followed by melting curve to screen for *TP53* mutations.

The peptide nucleic acids (PNAs) for the three different *TP53* mutations were designed using PNA BIO online programme to block the *TP53* wild-type allele. The PNAs specificity in blocking wild-type alleles was validated through qPCR for 40 cycles with the general probe using HGD. The threshold was equal to 0.01. The PNAs were partially efficient in blocking the wild-type allele in HGD, as the CT means were delayed by more than ten cycles with PNA compared to without PNA (Figure 4.11A; Table 4.6). Moreover, PNAs were specific to block *TP53* wild-type allele more than mutant allele as CT means of the homozygous cell lines DNA (HT29) was only delay by almost one cycle with PNA compared to without PNA (Figure 4.11B; Table 4.6).

TP53	**DNA**	**CT mean (General probe)**	
		No PNA	**with PNA**
c.818G>A	HGD	27.3	37.3
c.524G>A	HGD	25.6	36
c.743G>A	HGD	26	39.7

Table 4.6. CT means of *TP53* wild-type amplicons with and without PNA. CT means of *TP53* (c.818G>A, c.524G>A and c.743G>A) in HGD, through qPCR using a general probe with and without PNA.

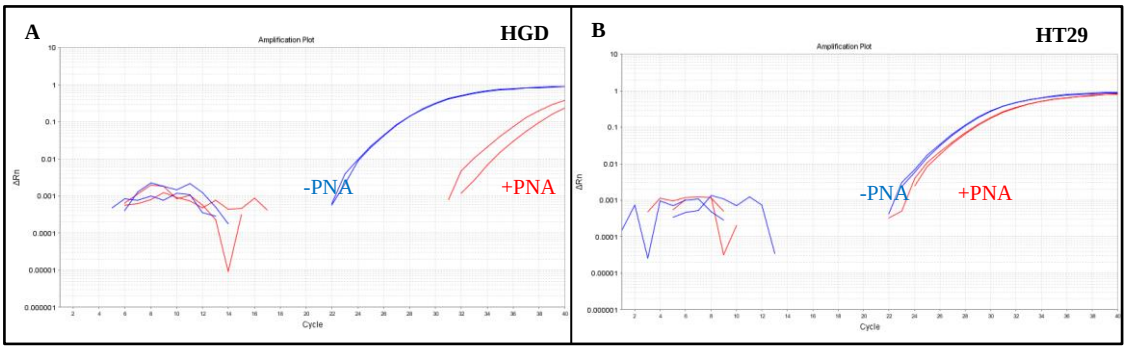


Figure 4.11. *TP53* mutation detection with and without PNA. *TP53* wild-type (HGD) with and without PNA (A) and *TP53* mutation in homozygous cell line (HT29) (B), using qPCR and general probe. The blue line represent amplification without PNA and the red line with PNA. Threshold=0.01.

The validation of HRMC was carried out first in HGD, cell line DNA with *TP53* mutation and five different cell line DNAs with wild-type *TP53* to set the threshold for the significant relative signal differences between any DNA and the reference wild-type HGD. The relative signal differences of the amplicon, which consist of a *TP53* (c.818G>A) or (c.524G>A) mutant allele were revealed by negative values (Figure 4.12A & B). On the other hand, the relative signal differences of the amplicon which consist of *TP53* (c.743G>A) mutant allele had positive values (Figure 4.12C).

By comparing the relative signal differences between 5 different cell lines DNAs with *TP53* wild-type alleles and reference DNA (HGD), the differences was more than -8 in the region of *TP53* (c.818G>A), more than -4 in the region of *TP53* (c.524G>A) and less than 1.5 in the region of *TP53* (c.743G>A) (Figure 4.12A, B & C). Therefore, the presence of *TP53* (c.818G>A) mutation was detected with relative signal differences ≤ -8 and for *TP53* (c.524G>A) mutation was ≤ -4. While, the presence of *TP53* (c.743G>A) mutation was detected with relative signal differences ≥ 1.5.

The H23 cell line DNA was examined as it consists of *TP53* (c.738G>C) mutation (www.cancer.sanger.ac.uk/COSMIC) which located in the PNA blocking region of *TP53* (c.743G>A). The results confirmed the presences of *TP53* mutations in H23 cell line DNA as the relative signal differences were equal to 11 (Figure 4.12C).

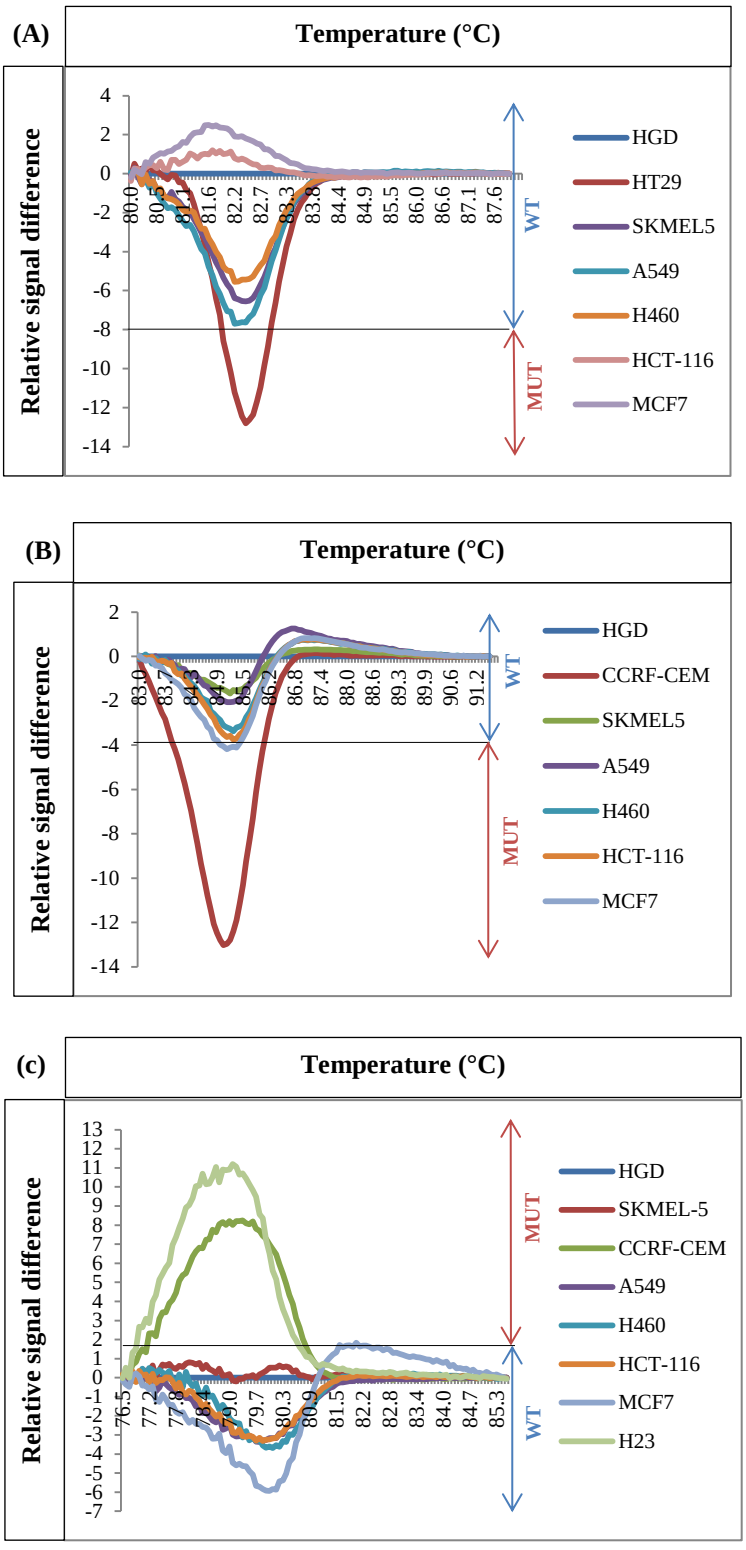


Figure 4.12. Relative signal differences between different cell line DNA and HGD. Relative signal differences (melting curves) between cell line DNAs for: (A) *TP53* (c.818G>A), (B) *TP53* (c.524G>A) or (C) *TP53* (c.743G>A) mutant allele and reference DNA, which consist of *TP53* wild-type alleles (HGD). The threshold line is at -8 of relative signal differences for *TP53* (c.818G>A), -4 for *TP53* (c.524G>A) and +1.5 for (c.743G>A). MUT; represent mutant alleles and WT; represents wild-type alleles.

The validation of HRMC with and without PNA in detecting *TP53* mutations was carried out using HGD as a reference for wild-type allele and dilution of HT29 cell line DNAs with HGD (50%, 10%, 1% and 0.1% cell line DNA). The HRMC with PNA was efficient in detecting up to 1% of *TP53* (c.818G>A) mutation fraction compared to 50% mutation fraction without PNA, with significant relative signal differences ≤ -8 degrees (Figure 4.13A & B).

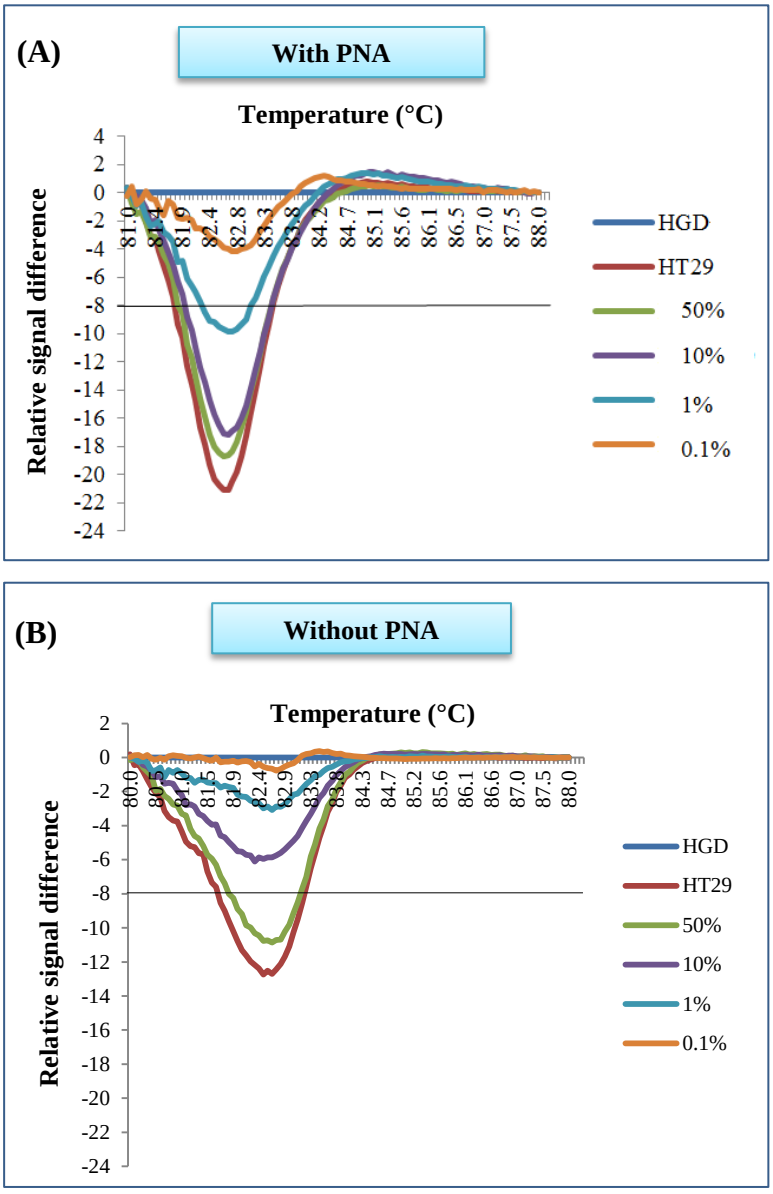


Figure 4.13. Relative signal differences between diluted cell line DNA and HGD with and without PNA. Relative signal differences (melting curves) between dilution of HT29 cell line DNA (*TP53* (c.818G>A)) and reference DNA which consist of *TP53* wild-type allele (HGD) with (A) and without DNA (B). The threshold line is at -8 relative signal differences for *TP53* (c.818G>A).

The validation of HRMC with PNA in detecting *TP53* mutations was carried out using HGD as a reference for wild-type allele, cell lines DNA with *TP53* mutations and dilution of cell line DNAs with HGD (50%, 10%, 1% and 0.1% cell line DNA) (Figure 4.14A, B & C).

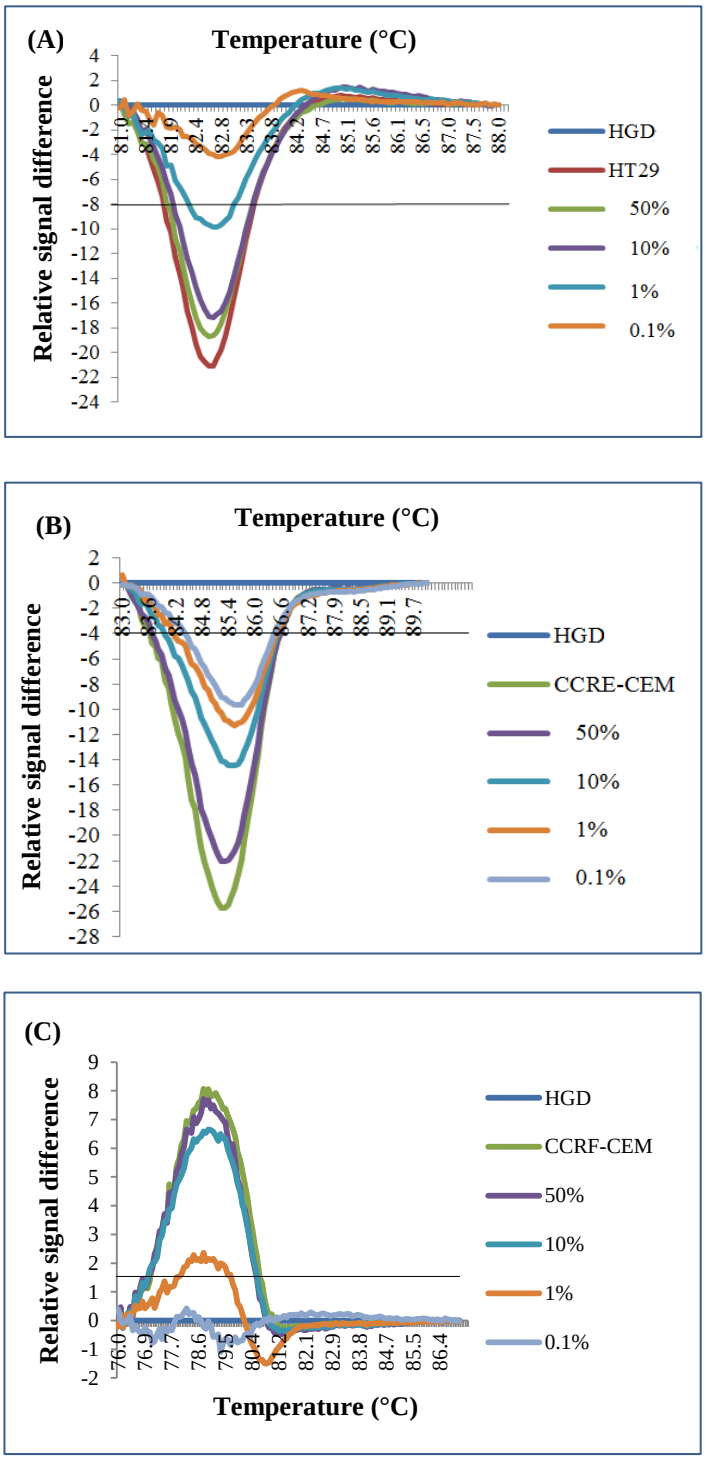


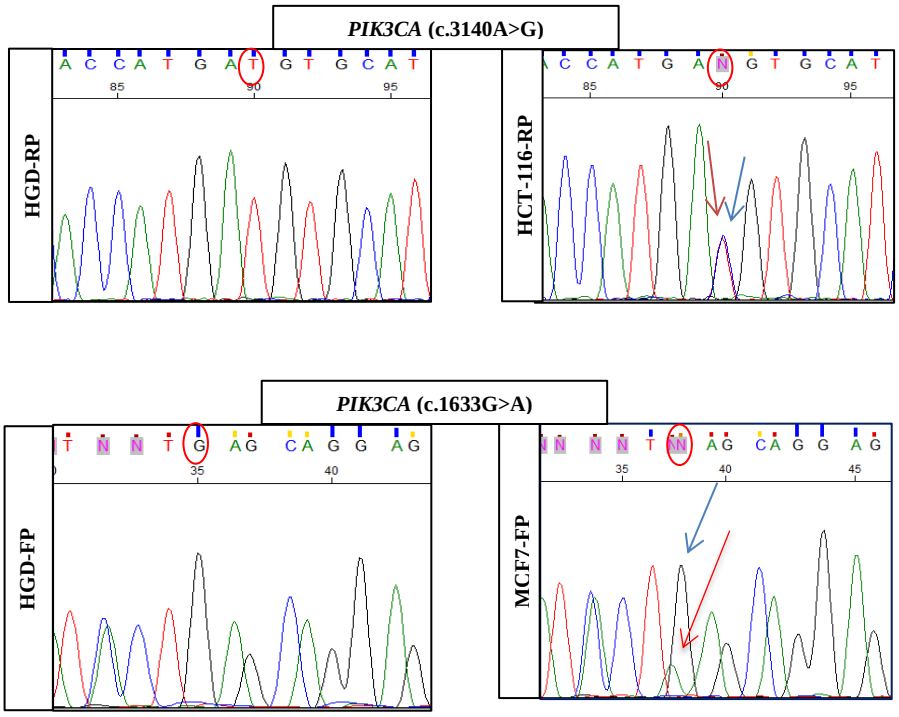
Figure 4.14. Relative signal differences between diluted cell line DNA and HGD. Relative signal differences (melting curves) between dilution of the positive cell line DNA with: (A) *TP53* (c.818G>A), (B) *TP53* (c.524G>A) and (C) *TP53* (c.743G>A) mutant alleles and

reference DNA which consist of *TP53* wild-type allele (HGD). The threshold line is at -8 relative signal differences for *TP53* (c.818G>A), -4 for *TP53* (c.524G>A) and +1.5 for *TP53* (c.743G>A).

The HRMC with PNA was efficient in detecting up to 0.1% of *TP53* (c.524G>A) mutation with relative signal differences ≤ -4 degrees (Figure 4.14B). Also, the HRMC with PNA was efficient in detecting up to 1% of *TP53* (c.818G>A) and (c.743G>A) mutations with relative signal differences ≤ -8 and ≥ +1.5 degrees respectively (Figure 4.14A & C).

4.2.1.4 DNA Sanger sequencing

Sanger sequencing was used to confirm detection of *PIK3CA* and *TP53* mutations in cell line DNA. The results confirmed the presence of *PIK3CA* (c.3140A>G) mutant allele in HCT-116 cell line DNA, *PIK3CA* (c.1633G>A) mutant allele in MCF7 cell line DNA, *PIK3CA* (c.1624G>A) mutant allele in SW948 cell line DNA, *TP53* (c.743G>A) and (c.524G>A) mutant alleles in CCRF-CEM cell line DNA and *TP53* (c.818G>A) mutant allele in HT29 cell line DNA (Figure 4.15).



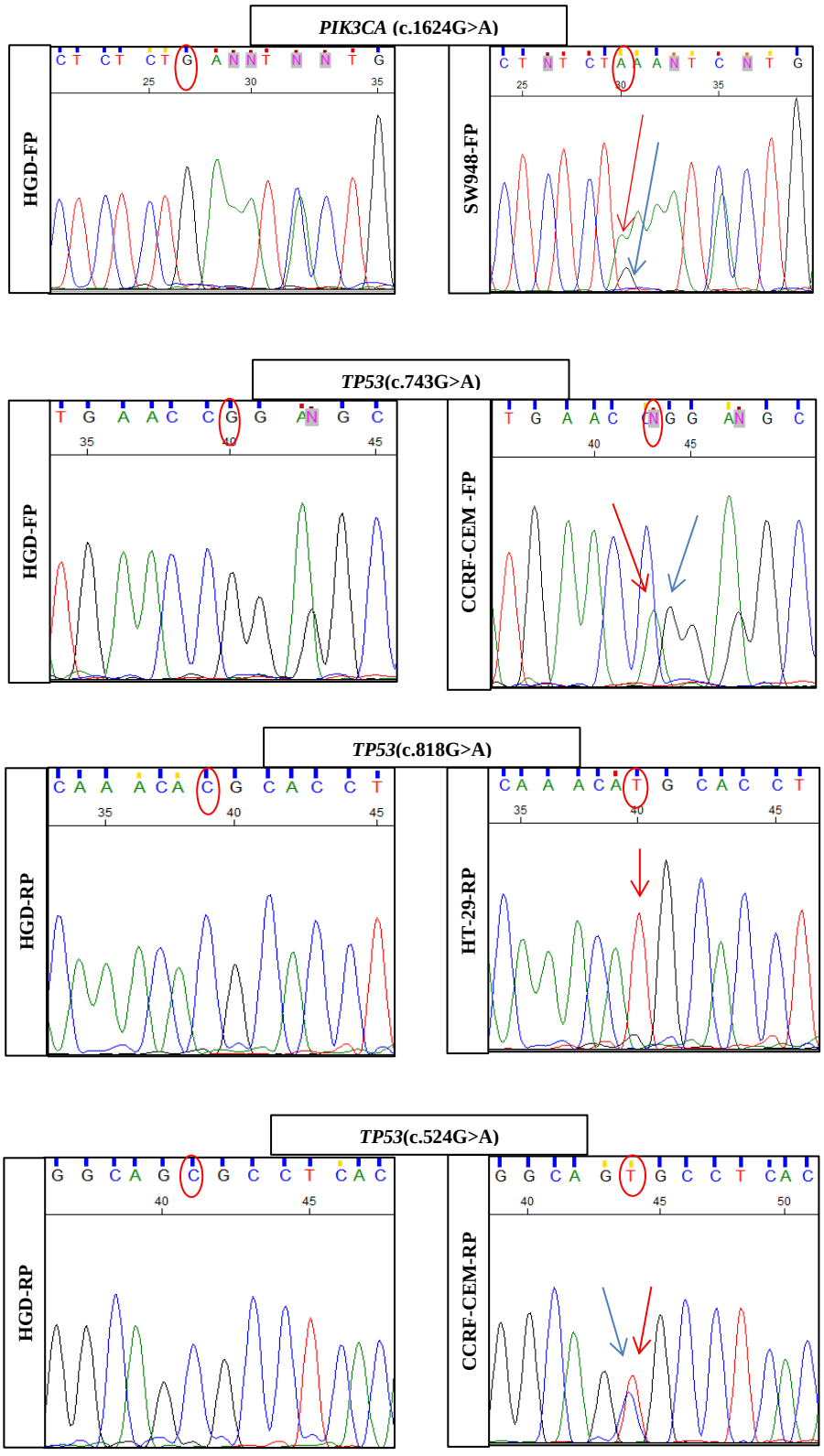


Figure 4.15. Sanger Sequencing results of HGD and cell line DNA with *PIK3CA* or *TP53* mutations. Sanger sequencing results of HGD and different cell lines DNA with *PIK3CA* or *TP53* mutations; HCT-116 (*PIK3CA*-c.3140A>G), MCF7 (PIK3CA-c.1633G>A), SW948 (*PIK3CA*-c.1624G>A), *TP53* (c.818G>A), *TP53* (c.524G>A) and *TP53* (c.743G>A). Blue arrows pointed to wild-type alleles and red arrows to mutant alleles. FP; indicates forward primer, RP; indicates the reverse primer reading sequence and circled base represent the wild-type or mutant allele.

The presence of different *PIK3CA* or *TP53* hotspots mutations and *PIK3CA* CNAs in all the chosen cell lines based on COSMIC data (www.sanger.ac.uk/COSMIC) were confirmed by different methods (Table 4.7).

Cell line	Mutation or CNA	ddPCR	qPCR	HRMC	Sanger sequencing
HCT-116	***PIK3CA* (c.3140A>G)**	✓	✓	-	✓
MCF7	***PIK3CA* (c.1633G>A)**	✓	✓	-	✓
SW948	***PIK3CA* (c.1624G>A)**	✓	✓	-	✓
CCRF-CEM	***TP53* (c.524G>A)**	✓	-	✓	✓
CCRF-CEM	***TP53* (c.743G>A)**	✓	-	✓	✓
HT29	***TP53* (c.818G>A)**	✓	-	✓	✓
RERF	***PIK3CA* CNA**	✓	-	-	-

Table 4.7. Detection of somatic mutations in cell lines DNAs. Different *PIK3CA* and *TP53* mutations or *PIK3CA* CNAs detection in Cell line DNAs with different methods.

4.2.2 Examination of Neocent tissue DNA for *PIK3CA* or *TP53* mutations using different methods

After the validation of *PIK3CA* and *TP53* mutation assays in cell lines DNA, 26 Neocent breast cancer tissues DNAs were examined for the presence of *PIK3CA* or *TP53* mutations with the previous different methods.

4.2.2.1 Examination of Neocent tissue DNA for *PIK3CA* or *TP53* mutations using ddPCR

After the validation of ddPCR with the designed assays in detecting *PIK3CA* and *TP53* mutations in diluted cell line DNAs, 26 Neocent tissues DNA were examined for the presence of *PIK3CA* or *TP53* mutations using ddPCR.

The events results, which are the number of positive droplets with mutant or wild-type alleles in each DNA sample, were considered as positive for *PIK3CA* mutation with ≥ 3 positive droplets with mutant allele. Out of 26 Neocent tissue DNAs, 8 DNA samples were positive for *PIK3CA* (c.3140A>G), 2 DNA samples for *PIK3CA* (c.1624AG>A), 4 DNA samples for *TP53* (c.743G>A) and one DNA sample for *TP53* (c.524G>A) mutations (Figure 4.16A, B, C & D; Table 8). None of 26 Neocent tissue DNAs was positive for *PIK3CA* (c.1633G>A) or *TP53* (c.818G>A) mutations. Although the

template concentrations of different DNA samples were equal (10 ng/3.6 µL), there was a variation in the number of droplets with wild-type alleles in different DNA samples (Figure 4.16), and that could be due to sampling quantification or template quality variations.

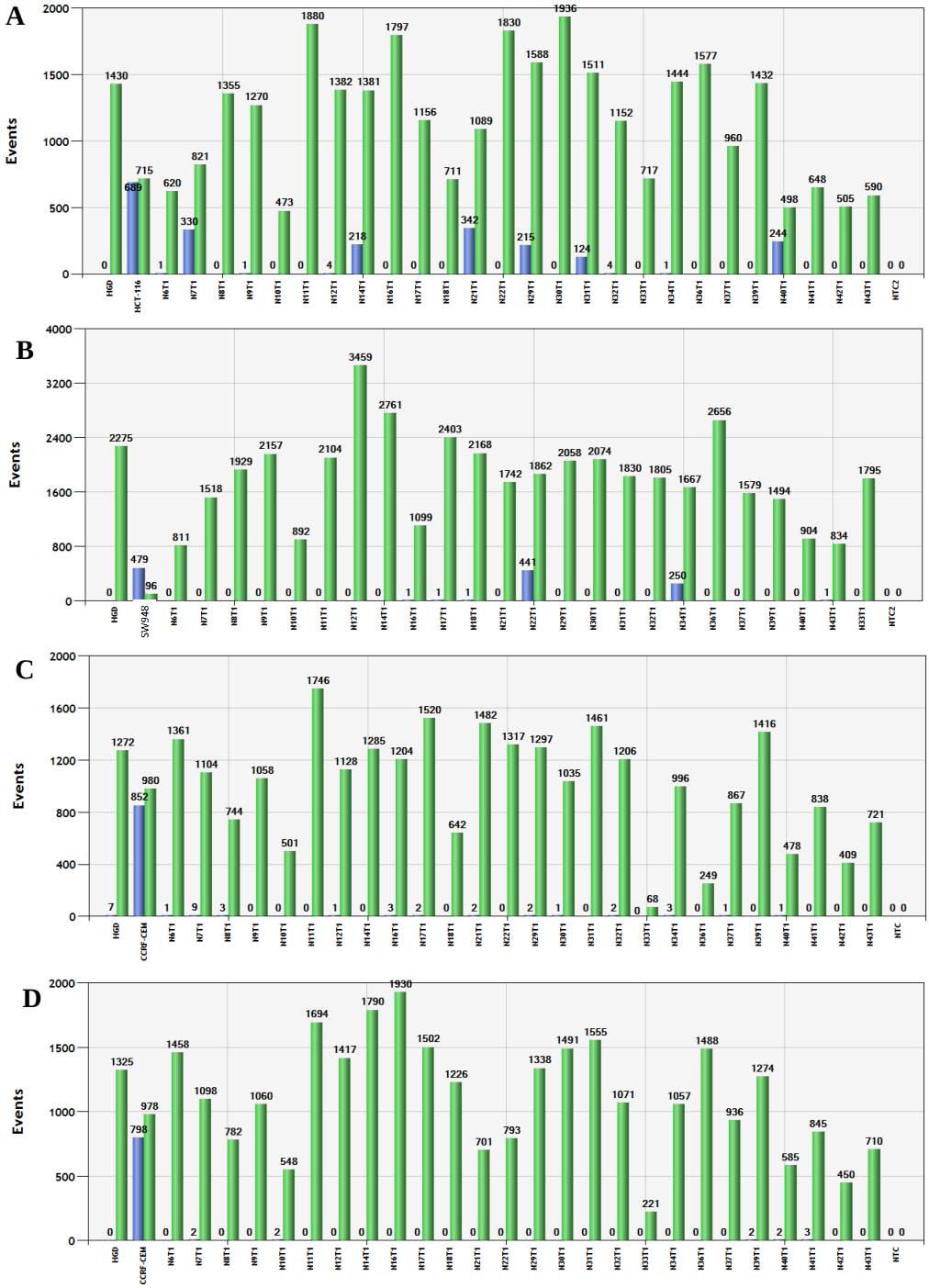


Figure 4.16. The ddPCR event results for *PIK3CA* or *TP53* mutations in 26 Neocent tissues DNA. Column graphs show the events which are the number of positive wild-type allele droplets (green column) and mutant allele droplets (blue column). (A) *PIK3CA* (c.3140A>G), (B) *PIK3CA* (c. 1624G>A), (C) *TP53* (c.743G>A) and (D) *TP53* (c.524G>A).

The fractional abundance % (which is the percentage of droplets number with mutant allele to total number of droplets with mutant and wild-type alleles) for all *PIK3CA* mutant mutation were more than 5% (range between 7.2% and 32.6%) except for N12T1 and N32T1 DNA samples were less than 1% (Table 4.8). On the other hand, the fractional abundance % for all *TP53* mutant positive results was less than 1% (Table 4.8).

The concentrations of *PIK3CA* mutations were more than 10 copies/ µL (range between 11.9 and 40.7 copies/ µL) except for N12T1 and N32T1 DNA samples were less than 1 copies/ µL (Table 4.8). While all *TP53* mutations concentrations was less than 1 copies/ µL (Table 4.8).

***PIK3CA* mutations**	**Neocent tissue DNA**		**Events of mutant allele**		**Fractional abundance (%)**	**Concentrations (copies/µL)**	
			MUT	WT		MUT	WT
***PIK3CA* (3140A>G)**	N7	T1	330	821	28.3	30.8	78
		T3	554	1353	28.4	50.6	128
	N12	T1	4	1382	0.12	0.6	208
		T3	0	1866	0	0	298
	N14T1		218	1381	12.8	28.7	195
	N21T1		342	1089	23.2	40.7	135
	N29T1		215	1588	11.3	21.8	172
	N31T1		124	1511	7.2	11.9	154
	N32T1		4	1152	0.33	0.4	122
	N40	T1	244	498	32.6	29.1	60
		T3	373	1464	19.7	29.9	122
***PIK3CA* (1624G>A)**	N22T1		441	1862	18	50.2	228
	N34T1		250	1667	12.2	29.5	212
***TP53* (c.743G>A)**	N7	T1	9	1104	0.8	0.8	108
		T3	0	1355	0	0	245
	N8	T1	3	744	0.39	0.31	80
		T3	0	1191	0	0	201
	N16	T1	3	1204	0.23	0.15	215
		T3	2	1296	0.14	0.23	159
	N34T1		3	996	0.28	0.4	158
***TP53* (c.524G>A)**	N41T1		3	845	0.34	0.36	105

Table 4.8. The ddPCR results of Neocent tissue's DNA with *PIK3CA* mutations. Fractional abundance (%), events and the concentrations (copies/µL) of mutant and wild-type allele of *PIK3CA* (3140A>G), *PIK3CA* (1624G>A), *TP53* (c.743G>A) and *TP53* (c.524G>A) mutations in positive Neocent tissue DNA, using ddPCR.

Some of Neocent tissues DNAs after completing treatment (T3) which were matched with the baseline DNAs (T1) with *PIK3CA* or *TP53* mutations were examined for the presence of these mutations using ddPCR. Three DNA samples after completing

treatment (T3) were tested for the presence of *PIK3CA* (c.3140A>G) or *TP53* (c.743G>A) mutations. Two out of three DNA samples after completing treatment (T3) were positive for *PIK3CA* (c.3140A>G) mutation, while none of three samples had (c.743G>A) mutations (Figure 4.17; Table 4.8).

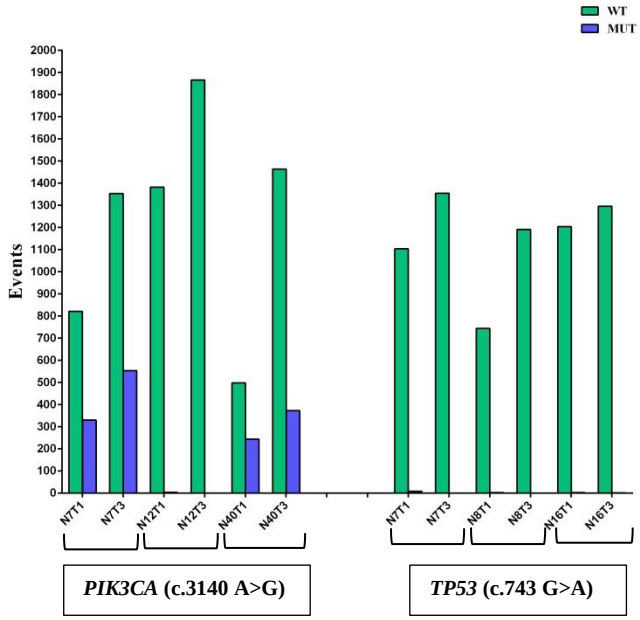


Figure 4.17. The ddPCR event results for *PIK3CA* (c.3140A>G) and *TP53* (c.743G>A) mutations in Neocent tissues DNA before (T1) and after completing treatment (T3). Column graphs show the events which are the number of positive wild allele droplets (green column) and mutant allele droplets (blue column) of *PIK3CA* (3140G>A) and *TP53* (c.743G>A) mutations in Neocent tissue DNAs before treatment (T1) and their matched after completing treatment tissue DNAs (T3).

The previous validated primers and probes were used through ddPCR to detect the presence of *PIK3CA* CNA in 26 Neocent tissues DNA. Two out of 26 DNA samples (N8T1 and N11T1) consist of CNA within *PIK3CA* gene as their *PIK3CA* copy number was >2.2 and their ratio with the copy number of *RPPH1* gene were > 1.2 (Figure 4.18A & B; Table 4.9). On the other hand, 2 Neocent tissues DNA (N36T1 and N37T1) had low copy number of *PIK3CA* gene which equal to 1.55 and 1.42 respectively and their ratios were equal to 0.77 and 0.71 respectively (Figure 4.18A & B; Table 4.9). These finding could indicate deletion of *PIK3CA* gene in these 2 DNAs samples. Also, there were few DNA samples with high copy number (>2.2) or low copy number (<1.8) but their ratio with the copy number of *RPPH1* gene were within normal range (0.8 to 1.2) (Figure 4.18A & B).

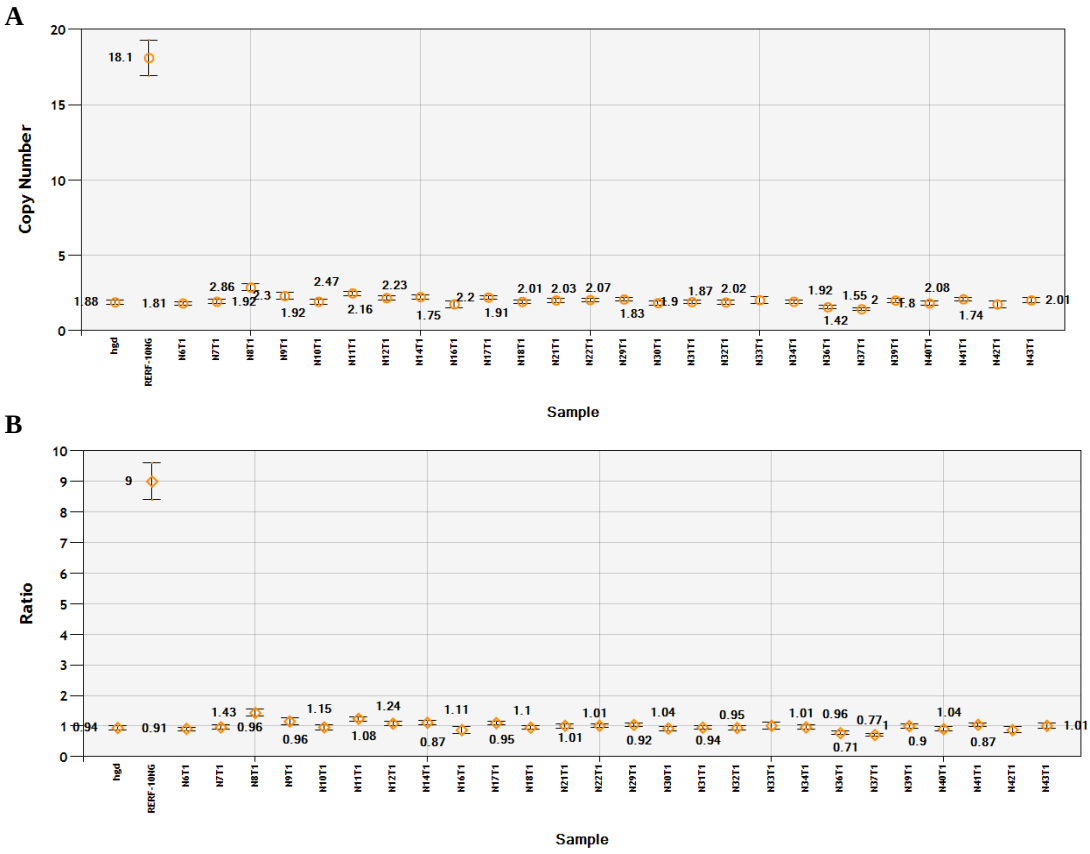


Figure 4.18. The ddPCR results of *PIK3CA* copy number aberrations (CNAs) in Neocent cancer tissues DNA. The copy number of *PIK3CA* gene (A) and the Ratio between *PIK3CA* gene copies to *RPPH1* gene copies (B) in 26 Neocent tissues DNA, using the ddPCR method.

Neocent tissue DNA	ddPCR		*PIK3CA* gene
	PIK3CA copy number	Ratio	
N8T1	2.86	1.43	**Gained**
N11T1	2.47	1.24	**Gained**
N36T1	1.55	0.77	**Deletion**
N37T1	1.42	0.71	**Deletion**

Table 4.9. The ddPCR results of *PIK3CA* copy number aberrations (CNAs) in Neocent breast cancer tissues DNA. The copy number of *PIK3CA* gene and its ratio to the copy number of *RPPH1* gene in Neocent tissues DNA, using the ddPCR method.

4.2.2.2 Examination of Neocent tissue DNA for *PIK3CA* mutations using qPCR with PNA

After the validation of qPCR with PNA and *PIK3CA* mutation assays in cell line DNA and 10 Neocent tissues and matched plasma DNA (Pilot study), 26 Neocent tissue DNAs were examined for the presence of *PIK3CA* mutation through qPCR with PNA

for 50 cycles. The *PIK3CA* (c.3140A>G) mutation was detected in 7 out of 26 Neocent tissue DNAs and *PIK3CA* (c.1624G>A) mutation in 2 DNA samples (Figure 4.19). The *PIK3CA* (c.1633G>A) mutation wasn't detected in any of 26 Neocent tissue DNAs (CT mean >45), similar to lymphocyte DNA results (Figure 4.19).

The paired lymphocyte DNA of the Neocent tissue DNA with *PIK3CA* mutation was examined through qPCR with PNA for the presence of *PIK3CA* mutation to confirm somatic mutations. All the lymphocyte DNAs were negative for the *PIK3CA* mutation (CT=50). Also, two of after treatment Neocent tissue DNA (N7T3 and N40T3) were examined through qPCR with PNA as they were matched for baseline DNA (T1) with *PIK3CA* mutation. The *PIK3CA* (c.3140A>G) mutation was detected in these two after treatment DNAs (T3) (with CT mean <45) (Figure 4.19; Table 4.10).

***PIK3CA* mutation**	**Neocent tissue DNA**	**CT Mean (MUT) with PNA**
***PIK3CA*-c.3140A>G**	N7 T1	30.3
	T3	27.4
	N14T1	30.7
	N21T1	29.1
	N29T1	34.1
	N31T1	31.1
	N32T1	40.6
	N40 T1	33.8
	T3	28.1
***PIK3CA*-c.1624G>A**	N22T1	30.6
	N34T1	31.7

Table 4.10. Neocent tissue's DNAs with *PIK3CA* mutations. CT means of *PIK3CA* mutant allele (MUT) in Neocent tissues DNA and their available matched after treatment tissues DNA (T3), using qPCR with PNA and duplex reaction (mutant and wild-type probes).

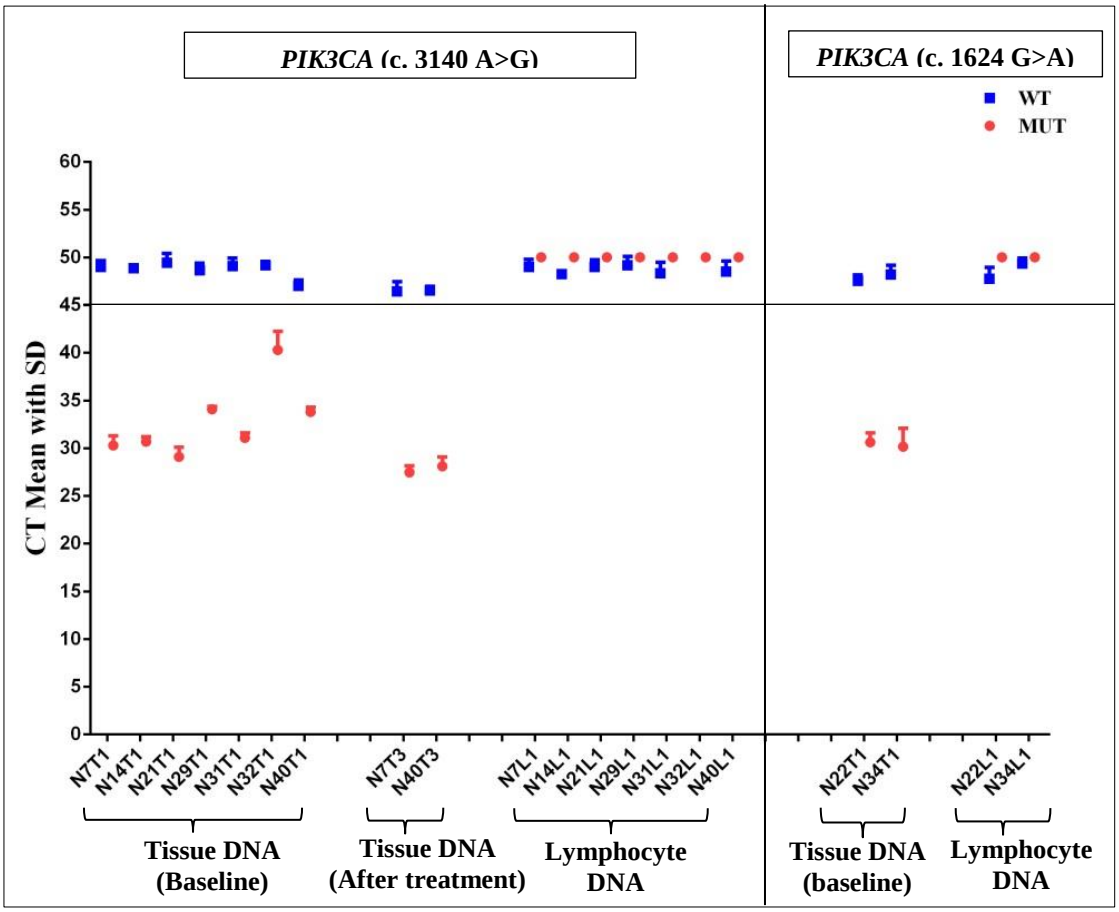


Figure 4.19. The qPCR results of *PIK3CA* gene in Neocent breast cancer patients. *PIK3CA* wild-type and mutant alleles examination in 26 Neocent tissues DNA using qPCR. Dotted graphs show CT mean with standard deviation (SD) of *PIK3CA* (c.3140A>G) and (c.1624 G>A) in Neocent tissues DNA (T1), matched after treatment tissues DNA (T3) and matched lymphocytes, using qPCR with PNA and duplex reaction (mutant and wild-type probes). Black line represents the CT mean threshold= 45 (CT mean <45 indicate the amplification). Blue squares represent wild-type alleles (WT), and red circles represent mutant alleles (MUT).

4.2.2.3 Examination of Neocent tissue DNA for *TP53* mutations using HRMC with PNA

After the validation of HRMC with PNA and *TP53* assays in detecting *TP53* mutations, they were used to examine 26 Neocent tissue DNAs and their paired lymphocyte for the presence of *TP53* mutations. HGD was used as a reference for the *TP53* wild-type. The result showed 5 out of 26 Neocent tissues DNA were positive for *TP53* (c.818G>A) and 10 were positive for *TP53* (c.524G>A) (Appendix 1 & 2; Table 4.11). However, there were variations in the positive results as some of the matched lymphocyte DNA was also considered as positive for *TP53* mutation according to the chosen threshold (e.g.,

N37L1 and N42L1) (Appendix 2). Also, one tissue DNA sample (N39T1) has higher melting curve than HGD (Appendix 2). Therefore, the paired lymphocyte DNAs was used as a reference to compare the melting curve between Neocent tissue DNA and the matched lymphocyte DNA with the same chosen threshold. The results showed 4 out of 26 Neocent tissues DNA were positive for *TP53* (c.818G>A), and 8 were positive for *TP53* (c.524G>A) (Appendix 1 & 2; Table 4.11). They were the same DNA samples which were positive for *TP53* mutation when HGD was used as a reference.

***TP53* mutation**	**HGD as a reference**	**Matched lymphocyte DNA as a reference**
***TP53* (c.818G>A)**	N9T1, N10T1*, N32T1, N33T1 and N41T1	N9T1, N32T1, N33T1 and N41T1
***TP53* (c.524G>A)**	N6T1, N9T1, N10T1, N21T1, N33T1, N37T1*, N39T1, N40T1, N41T1 and N42T1*.	N6T1, N9T1, N10T1, N21T1, N33T1, N39T1, N40T1 and N41T1.

Table 4.11. Neocent tissue's DNAs with *TP53* mutations. Neocent tissue DNAs which are positive for *TP53* (c.818G>A) and (c.524G>A) mutations with HGD or matched lymphocyte as references, using HRMC with PNA. The Neocent samples* were positive for *TP53* mutation with HGD as a reference, but not with matched lymphocyte as a reference.

The examination of *TP53* (c.52G>A) and (c.818G>A) mutations in Neocent tissues DNA were carried out as previously described. However, the DNAs with *TP53* mutation showed multiple variations in the melting curve, when the experiments were repeated. Moreover, the melting curves of some Neocent tissue DNAs for *TP53* (c.743G>A) were contain two peaks, even with repeating the experiment (Figure 4.20), which were different than the melting curve of the cell line DNA (Figure 4.14C). Moreover, it works with cell line models [*TP53* (c.743G>A) positive cell line (CCRF-CEM) and negative cell line (HT29)], but it was highly variable with Neocent tissue samples, in the same experiment (Figure 4.20).

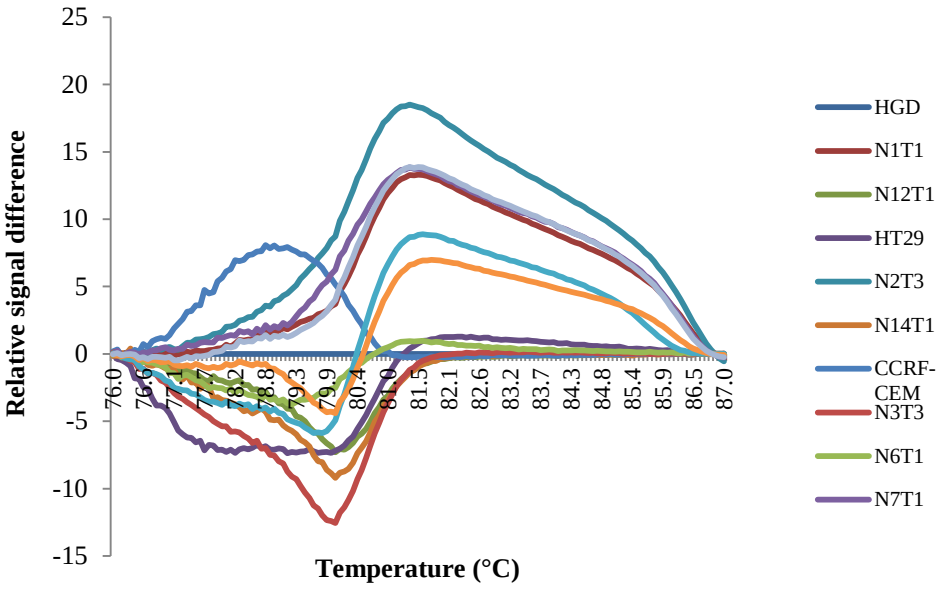


Figure 4.20. Melting curve differences of *TP53* (c.743G>A) between Neocent tissue DNA and HGD. The melting temperature differences (melt curve) between 10 Neocent tissues DNA and HGD. It also showed the melt curve of *TP53* (c.743G>A) positive cell line (CCRF-CEM) and negative cell line (HT29).

4.2.2.4 Examination of Neocent tissue DNA for *PIK3CA* or *TP53* mutations using DNA Sanger sequencing

All the Neocent tissues DNAs at baseline (T1), which showed the presence of *PIK3CA* or *TP53* mutations by ddPCR, qPCR with PNA or HRMC with PNA were sequenced using DNA Sanger sequencing.

Based on DNA Sanger sequencing; the *PIK3CA* (c.3140A>G) mutant allele was detected in 6 out of 8 Neocent tissue DNAs and *PIK3CA* (c.1624G>A) in 2 out of 2 Neocent tissue DNAs (Figure 21). On the other hand, none out of 19 Neocent tissues DNA detect the presence of any of *TP53* mutations, which detected with HRMC (n=15) and ddPCR (n=4). But, the *TP53* (c.817C>T) mutant base was detected in N9T1 DNA using Sanger sequencing (Figure 22) and this mutation was reported in COSMIC data (www.cancer.sanger.ac.uk/COSMIC).

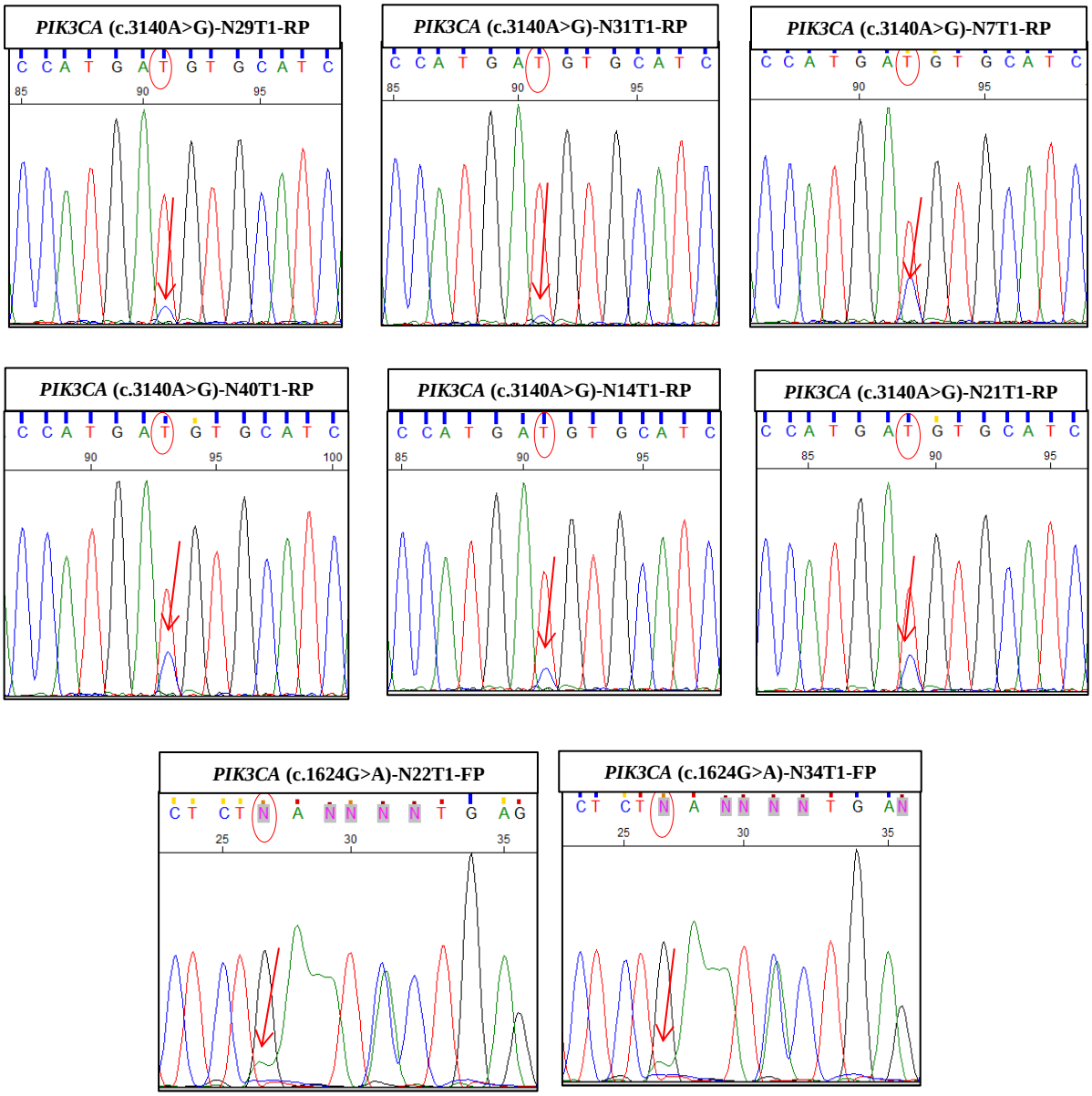


Figure 4.21. DNA Sanger sequencing results of Neocent tissues DNA with *PIK3CA* (c. 3140A>G & c.1624G>A) mutations. Red arrows pointed to the mutant alleles, and the circle base represents *TP53* wild-type or mutant alleles. RP; indicates the reverse primer reading sequence.

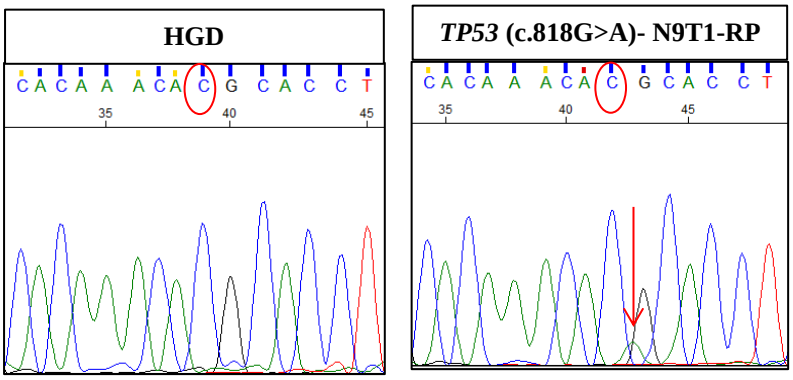


Figure 4.22. DNA Sanger sequencing results of *TP53* (c.818G>A) region of HGD and N9T1 DNA. The circle base is the *TP53* wild-type base, and the red arrow pointed to *TP53* (c.817C>T mutation). RP; indicates the reverse primer reading sequence.

4.2.3 Correlation between the presence of *PIK3CA* or/and *TP53* mutations and response to the neoadjuvant treatment.

7 out of 10 (70%) of the Neocent patient with *PIK3CA* mutation had significant radiological response regardless of treatment received (4 of them had endocrine therapy, and 3 of them had chemotherapy). 3 out of 4 (75%) of Neocent patients with *PIK3CA* mutation and received chemotherapy had a significant radiological response. Also, 4 out of 6 (66.7%) of Neocent patients with *PIK3CA* mutation and received endocrine therapy had significant radiological response. On the other hand, 6 out of 13 (46.2%) of a Neocent patient with *TP53* mutation and 100% of *PIK3CA* CNAs had significant radiological response regardless of treatment received. 3 out of 6 (50%) of a Neocent patient with *TP53* mutation and received chemotherapy had a significant radiological response and 3 out of 7 (42.9%) of patient with *TP53* mutation and received endocrine therapy had a significant radiological response (Table 12).

5 out of 26 (19.2%) of Neocent breast cancer tissue has both *PIK3CA* and *TP53* mutations and 3 out of 26 (11.5%) harboured two different *TP53* mutations. 10 out of 26 (38.5%) Neocent patients didn't respond to either chemotherapy or endocrine therapy. 4 out of 10 (40%) of non-responder Neocent tissues DNA harbour a combination of *PIK3CA* and *TP53* mutations or 2 different *TP53* mutations. 2 out of 3 (66.7%) Neocent tissue DNA with *PIK3CA* mutation and didn't respond to treatment harbour a combination of *PIK3CA* and *TP53* mutations and 4 out of 5 (80%) of Neocent

tissue DNA with *TP53* mutation and didn't respond to treatment harbour either a combination of *PIK3CA* and *TP53* mutations or 2 different *TP53* mutations (Table 4.12).

Neocent tissue DNA	**Detection methods**					**Treatment**	**Response**
	PCR with PNA		ddPCR		DNA Sanger sequencing		
	qPCR	HRMC	Somatic mutation	*PIK3CA* CNAs			
N6T1	-	*TP53* (524G>A)	-	Normal	-	Chemo	Unknown
N7T1* & T3	*PIK3CA* (3140A>G)	-	*PIK3CA* (3140A>G) & *TP53* (743G>A)	Normal	*PIK3CA* (3140A>G)	Endo	No
N8T1 &T3	-	-	*TP53* (743G>A)	Gain	-	Endo	Yes
N9T1*	-	*TP53* (524G>A) & *TP53* (818G>A)	-	Normal	-	Chemo	No
N10T1	-	*TP53* (524G>A)	-	Normal	-	Endo	No
N11T1	-	-	-	Gain	-	Chemo	Yes
N12T1 & T3	*PIK3CA* (3140A>G)	-	*PIK3CA* (3140A>G)	Normal		Endo	No
N14T1	*PIK3CA* (3140A>G)	-	*PIK3CA* (3140A>G)	Normal	*PIK3CA* (3140A>G)	Endo	Yes
N16T1 &T3	-	-	*TP53* (743G>A)	Normal	-	Chemo	Yes
N21T1*	*PIK3CA* (3140A>G)	*TP53* (524G>A)	*PIK3CA* (3140A>G)	Normal	*PIK3CA* (3140A>G)	Endo	Yes
N22T1	*PIK3CA* (1624G>A)		*PIK3CA* (1624G>A)	Normal	*PIK3CA* (1624G>A)	Chemo	Yes
N29T1	*PIK3CA* (3140A>G)		*PIK3CA* (3140A>G)	Normal	*PIK3CA* (3140A>G)	Endo	Yes
N31T1	*PIK3CA* (3140A>G)		*PIK3CA* (3140A>G)	Normal	*PIK3CA* (3140A>G)	Endo	Yes
N32T1*	*PIK3CA* (3140A>G)	*TP53* (818G>A)	*PIK3CA* (3140A>G)	Normal	*PIK3CA* (3140A>G)	Chemo	Yes
N33T1*	-	*TP53* (524G>A) & *TP53* (818G>A)	-	Normal	-	Endo	Yes
N34T1*	*PIK3CA* (1624G>A)	-	*PIK3CA* (1624G>A) & *TP53* (743G>A)	Normal	*PIK3CA* (1624G>A)	Chemo	No
N36T1	-	-	-	Deletion	-	Endo	Yes
N37T1	-	-	-	Deletion	-	Chemo	Yes
N39T1	-	*TP53* (524G>A)	-	Normal	-	Endo	Unknown
N40T1* &T3	*PIK3CA* (3140A>G)	*TP53* (524G>A)	*PIK3CA* (3140A>G)	Normal	*PIK3CA* (3140A>G)	Chemo	Yes
N41T1*	-	*TP53* (524G>A) & *TP53* (818G>A)	*TP53* (524G>A)	Normal	-	Endo	No

Table 4.12. Comparison of somatic mutations with different methods and their correlation with treatment response. Comparison of the results of different methods (qPCR, HRMC, ddPCR and DNA Sanger sequencing) in the detection of *PIK3CA* or *TP53* mutations or

PIK3CA CNAs in Neocent breast cancer tissues DNA. Also, treatment received by these patients and radiological response post-treatment. The Neocent samples* harbour a combination of *PIK3CA* and *TP53* mutations or 2 different *TP53* mutations.

Based on ddPCR results, the *PIK3CA* mutation and *TP53* mutation fraction reduced after neoadjuvant chemotherapy (N40T1, T3 and N16T1, T3 respectively) and associated with the radiological response. However, *PIK3CA* mutation fraction didn't change by neoadjuvant endocrine treatment (N7T1 &T3) and associated with no radiological response. While, *TP53* mutation fraction reduced after neoadjuvant endocrine therapy and associated with (N8T1, T3) or without radiological response (N7T1, T3), so a reduction in *TP53* mutation after neoadjuvant endocrine therapy may have no effect on the radiological response.

4.2.4 Whole exome sequencing

The whole exome sequencing was carried out on ten baseline Neocent tumour tissues DNAs (T1) and their matched tumour tissue after treatment (T3) (N7, N8, N9, N11, N12, N13, N16, N17, N18 and N30), by another research group. The bioinformatics analysis was done by Dr Jili Luo. The Neocent gene list is derived from the Neocent 9 pair exome dataset. The results showed the *TP53* and *PIK3CA* genes as common cancer driven genes in these ten Neocent tumour tissues DNAs (T1 and T3).

The whole exome sequencing results of ten Neocent tumour tissues DNAs were confirmed the presence of *PIK3CA* (c.3140A>G) mutation in N7T1 and T3 tissues DNAs (Table 4.13), which supporting our results provided from ddPCR, qPCR with PNA and Sanger sequencing. However, the whole exome sequencing didn't detect *TP53* hotspot mutations in N7T1, N8T1, N9T1 and N16T1, which was identified with HRMC or ddPCR. Also, it didn't confirm the presence of *PIK3CA* (c.3140A>G) mutation in N12T1, which was detected with ddPCR.

Neocent tumour tissues DNAs	**Somatic mutation**		
N7T1 & T3	*PIK3CA*	c.3140A>G	Nonsynonymous
N17T1 & T3	*PIK3CA*	c.3140A>T	Nonsynonymous
N8T1 & T3	*TP53*	c.49G>T	Nonsynonymous
N9T1 & T3	*TP53*	c.421C>T	Nonsynonymous
N11T1 & T3	*TP53*	c.32T>A	Nonsynonymous

Table 4.13. Somatic mutations in Neocent tissues using whole exome sequencing. *PIK3CA* and *TP53* mutations in the ten Neocent tissues DNAs from whole exome sequencing results.

4.2.5 Summary of *PIK3CA* and *TP53* somatic mutations analysis in breast cancer tissue and cf-DNA.

DNA extraction using Qiagen kit provide good quality DNA with higher concentration and better NGS results compared to phenol DNA extraction technique, and the pre-amplification of the examined area in DNA has no significant effect on ddPCR results compared to non-amplified DNAs.

The ddPCR method has higher sensitivity in detecting low-level *PIK3CA* or *TP53* mutations (up to 0.1% of mutant allele fraction) in cell line DNA, than that achieved with qPCR or HRMC with PNA and DNA Sanger sequencing. The qPCR with PNA and ddPCR assays showed similar sensitivity in detecting *PIK3CA* mutation in cell line DNA. However, the ddPCR has advantages over qPCR with PNA as it discovered additional mutations in breast cancer tissues not found by qPCR or Sanger sequencing and it can provide an accurate quantitative measurement of mutation.

The ddPCR was carried out to examine for *PIK3CA* mutation, *TP53* mutation and *PIK3CA* CNAs in Neocent tissue DNA (n=26) (Figure 4.23). The frequency of *PIK3CA* mutation in breast cancer tissue was slightly higher with ddPCR (10 out of 26; 38.5%) compared to qPCR with PNA (9 out of 26; 34.6%). While the frequency of *TP53* mutation was higher with HRMC (12 out of 26; 46.2%) compared to ddPCR (5 out of 26; 19.2%). However, there was a wide variation and failure in HRMC results which may cause false positive results as it didn't identify with ddPCR. The *PIK3CA* mutations were detected in after treatment (T3) Neocent tumour tissue DNA using qPCR with PNA (2 out of 2) and ddPCR (2 out of 3), but ddPCR couldn't detect the *TP53* mutation in after treatment (T3) Neocent tumour tissue DNA (0 out of 3). The frequency of *PIK3CA* somatic mutation was more frequent than the gain (2 out of 26; 7.7%) or deletion (2 out of 26; 7.7%) of gene copy number.

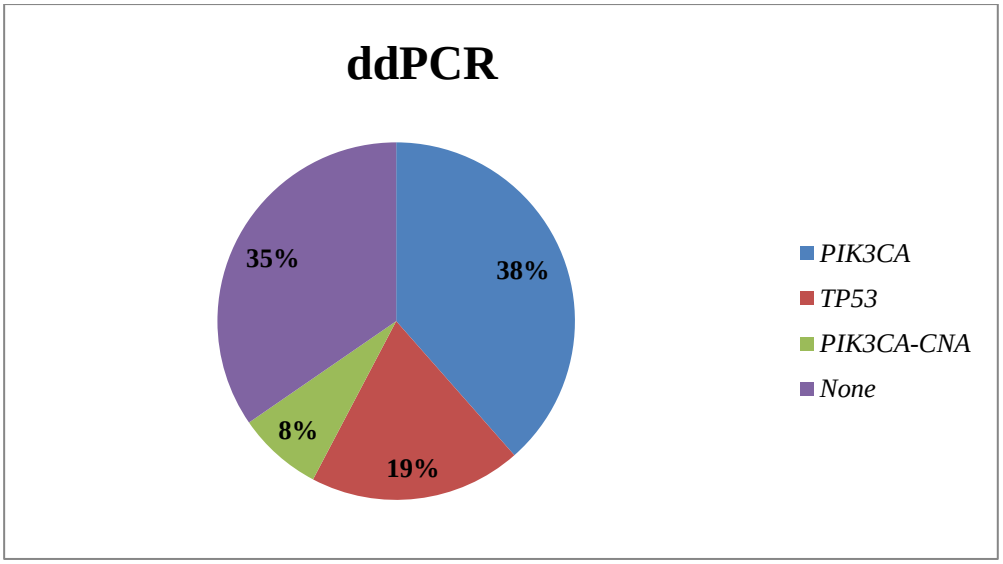


Figure 4.23. The frequencies of *PIK3CA* mutation, *TP53* mutation and *PIK3CA* CNAs in Neocent tissue (n=26), using the ddPCR.

The whole exome sequencing confirmed the presence of *PIK3CA* (c.3140G>A) mutation in N7T1 and T3, but it didn't detect *TP53* hotspot mutations in N7T1, N8T1, N9T1 and N16T1, which was detected with HRMC or ddPCR.

The presence of *PIK3CA* mutations or gain or deletion may associated with treatment response especially to chemotherapy. However, the presence of a combination of *PIK3CA* and *TP53* mutations or two different *TP53* mutations in the same Neocent breast cancer tissue correlate with treatment resistance to either chemotherapy or endocrine therapy which indicated by no radiological response. While, the presence of the *TP53* mutation may correlate with endocrine resistance.

4.3 Discussion

Tumour markers can help in the screening, diagnosis and prognosis of cancer, and provide important information for treatment needed in the context of personalised medicine. During recent years, many studies suggest different biomarker to improve identification of early-stage breast cancer patients, as serum proteins (Hao et al., 2016; Lee et al., 2016), metabolomics profiling (Barnes et al., 2014), plasma lipidomics profiling or surface proteins found on circulating extracellular vesicles (Moon et al., 2015). However, until now there are no molecular biomarkers for the early detection of breast cancer. Because, the biomarker results are difficult to replicate in different

studies for a combination of reasons, such as population and cancer heterogeneity, study sizes, the biological and experimental protocols variations.

The circulating cell-free tumour DNA (cf-DNA) as liquid biopsies can be considered to avoid partly the use of solid biopsies, as it is less invasive and described as more representative of tumour heterogeneity. However, cf-DNA can be found in blood in very low proportion, which is varying according to the nature and the progression of cancer. Fewer studies have been performed to detect cf-DNA in early cancer patients, due to its smaller amount, fragmentation and a short half-life (1–2 h) (Lo et al., 1999). For these reasons, detection of cf-DNA requires the use of highly sensitive, specific and ideally quantitative methods. Also, detecting low-level somatic mutations in cancer tissues DNA can be challenging in several settings: as in small biopsies, samples with significant inflammation, and in samples with low tumour content, due to the overwhelming presence of wild-type DNA.

Due to marked heterogeneity within breast cancer, no two tumours have the same molecular structure. Consequently, the most frequent mutations were examined to be biomarkers for detection of early breast cancer. Detection of *PIK3CA* and *TP53* mutations in cf-DNA with a highly sensitive assay is expected to be useful in the early detection of breast cancer, as *PIK3CA* and *TP53* are the most commonly mutated genes in patients with breast cancer (Stemke-Hale et al., 2008; Banerji et al., 2012; www.cancer.sanger.ac.uk/COSMIC) and play important roles in the biology of breast cancer. The *PIK3CA* gene, encoding the catalytic subunit of phosphoinositide 3-kinase (PI3K), which regulates cell survival, proliferation, motility, apoptosis and growth (Barbareschi et al., 2007). The tumour protein p53 is a tumour suppressor protein (encoded by *TP53* gene) and involved in the cell cycle, checkpoint control, repair of DNA damage, and apoptosis (Lehmann & Pietenpol, 2012). Several small studies have identified the presence of *PIK3CA* mutations in pre-invasive breast cancer, with the implication that these mutations are early events in breast cancer development (Miron et al., 2010; Li et al., 2010). Also, *TP53* mutation might be an early event occurring in DCIS or even earlier than DCIS stage (Ho et al., 2000; Done et al., 2001a; Done et al., 2001b; Kang et al., 2001; Wenjing et al., 2009). Several studies suggest that *PIK3CA* mutation may have predictive value for treatment response. However, other studies do not identify the significance of *PIK3CA* mutation in the prediction for treatment

response (Souglakos et al., 2009) or it may asoociated with treatment resistance (Mao et al., 2012).

Next generation sequencing (NGS) remains the gold standard for identifying DNA mutations, as it is reliable in distinguishing tumour from normal DNA or different tumour clones (Bai et al., 2014). Also, it provides accurate information of allele sequence and mutation frequency and detecting non-hotspot mutations (Gao et al., 2016). However, throughput and sensitivity issues for next-generation sequencing are gradually being addressed (Kinde et al., 2011; Milbury et al., 2012; Narayan et al., 2012; Black et al., 2015), so a good quality DNA need to be extracted for sequencing. Also, NGS remains a relatively expensive and time-consuming method. Therefore, mutation scanning is an attractive alternative that balances well resources versus the information obtained (Milbury et al., 2009). There are several methods for mutation scanning using PCR-based technology, including high-resolution melting (Herrmann et al., 2006), qPCR with peptide nucleic acid (PNA) and ddPCR. All of these methods are simple, cheap and easy to produce.

Results found in this thesis show that breast cancer tissues DNAs extracted using Qiagen kit have higher concentrations and better sequencing results with NGS, compared to DNAs extracted using phenol extraction technique. The PNAs were efficient in blocking the *PIK3CA* wild-type allele and increase the sensitivity of qPCR assays to detect *PIK3CA* mutation fraction up to 0.1% compared to 10%-50% without PNA. On the other hand, the PNAs were partially efficient in blocking the *TP53* wild-type allele, and that could be due to the high GC-rich regions, which may have an effect on PNA sensitivity.

The developed methods were optimised and shown to work via cell lines that contain the known mutation. The presence of *PIK3CA* and *TP53* mutations in these cell lines were confirmed by other studies, such as HCT-116 cell line was used as cell model for *PIK3CA* (c.3140A>G) mutation (Ligresti et al., 2009), MCF7 cell line proved the presence of *PIK3CA* (c.1633G>A) mutation (Hollestelle et al., 2007), SW948 cell line confirmed homozygous *PIK3CA* (c.1624G>A) mutation (Ahmed et al., 2013), HT-29 cell line confirmed to has *TP53* (c.818G>A) mutation (Bossi et al., 2006; Gurtner et al., 2010) and CCRF-CEM cell line proved a compound heterozygote for the *TP53*

mutations R175H (c.524G>A) and R248Q (c.743G>A) mutations (Cheng and Haas, 1990).

The sensitivity of the ddPCR assays to detect *TP53* mutations was 0.01-0.1% using cell lines DNA, and it is higher than HRMC sensitivity, which was 0.1-1%. On the other hand, the sensitivity of the qPCR with PNA and ddPCR assays to detect *PIK3CA* mutations proved to be 0.1 % using cell line DNA for both methods. Therefore, qPCR with PNA and ddPCR methods are equally sensitive to detect small *PIK3CA* mutation fraction up to 0.1% (10 pg mutant allele in 9.99ng wild-type allele). Although, the sensitivity of qPCR with PNA and ddPCR were equal using cell line DNA, the low proportion of *PIK3CA* mutation in Neocent tissues DNA (<1%) was more detected with ddPCR method compared to qPCR with PNA and that could be due to sample quality or experimental variability. While high *PIK3CA* mutant proportion (>5%) was detected by qPCR with PNA or ddPCR. Therefore, qPCR with PNA is a sensitive method for detecting small *PIK3CA* mutant proportion but may be affected by sample quality more than ddPCR, because ddPCR based on analysis of each droplet generated. Similarly, ddPCR can detect *PIK3CA* and *TP53* mutations in Neocent tissues DNA, but they couldn't be detected by Sanger sequencing as ddPCR can identify additional mutations in primary tumour tissues not found by Sanger sequencing (Beaver et al., 2014). This could be due to low detection sensitivity of Sanger sequencing, as it can only detect mutant sequences constituting more than 20% of the total genetic content (Kohlmann et al., 2011). Therefore, a significant proportion of *PIK3CA* or *TP53* mutations could be missed using Sanger sequencing due to a low proportion of tumour cells in the tumour tissue. Therefore, ddPCR is more sensitive in detecting low-level *PIK3CA* and *TP53* mutations compared to qPCR, HRMC with PNA and Sanger sequencing. This is supported by other researchers which they found the ddPCR is a highly sensitive (0.01%-0.1%) and specific method for the detection of *PIK3CA* mutant cf-DNA (Oshiro et al., 2015). Also, they found ddPCR enabled the detection of rare cf-DNA and thus its application to early cancer patients (Crowley et al., 2013; Diaz & Bardelli, 2014; Schwarzenbach, 2013). Thesis results using cell lines demonstrate that, pre-amplification of the specific sequences within the gene in DNA, which consists of mutation site, has no significant effect on detection of *PIK3CA* or *TP53* mutation using ddPCR compared to non-amplified DNAs.

Although ddPCR has higher sensitivity in detecting *TP53* mutations compared to HRMC with PNA, the frequency of *TP53* mutations in Neocent tissues was much higher (46.2%) with HRMC compared to ddPCR (19.2%). However, All *TP53* mutations detected by HRMC were not identified by ddPCR or Sanger sequence except one sample, which had a *TP53* (c.524G>A) by both HRMC and ddPCR. Moreover, HRMC had significant variable results with repeating experiments and failed melting curve with *TP53* (c.743G>A) which may suggest false positive results. The reason behind all of this variation with HRMC could be due to DNA quality and requires further experiment validation. The *TP53* mutations detected by HRMC or ddPCR couldn't be identify by Sanger sequencing, and that again could be due to low detection sensitivity of Sanger sequencing (Kohlmann et al., 2011) as all *TP53* mutant fractions were low (<1%) based on ddPCR results.

The frequency of *PIK3CA* mutation in Neocent tissue was slightly higher (38.5%) by ddPCR compared to qPCR with PNA (34.6%). All *PIK3CA* mutations detected by qPCR with PNA were also identified by ddPCR. But, ddPCR detected additional *PIK3CA* mutations in Neocent tissues not found by qPCR with PNA or Sanger sequencing. This supports by the findings of another research group, which demonstrates accurate mutation detection in tumour tissues using ddPCR (Beaver et al., 2014). The qPCR with PNA is a sensitive method for detecting small mutation proportion but may be affected by sample quality more than ddPCR, as ddPCR based on analysis of each droplet generated. Therefore, the previous evaluations of *PIK3CA* mutations in tumour tissues using qPCR with PNA might have underestimated the actual mutation frequency. The incidence of *PIK3CA* mutations found in Neocent tissue was within the range of previous reports which found somatic mutations in *PIK3CA* gene represent the most frequent alterations currently known in breast cancer patients, occurring at a frequency of 20–45% (Kalinsky et al., 2009; Dupont et al., 2011, Sudhakar et al., 2015).

80% of *PIK3CA* mutations detected by ddPCR were also detected by Sanger sequencing, but none for *TP53* mutations were detected by Sanger sequencing. That could be due to low proportion of tumour cells in the tumour tissue, which can vary widely from 20% to 95% (Cleator et al., 2006), and low detection sensitivity of Sanger sequencing (Kohlmann et al., 2011). Therefore, a significant proportion of *PIK3CA* or

TP53 mutations could be missed with Sanger sequencing due to a low proportion of tumour cells in the tumour tissue.

The *PIK3CA* and *TP53* mutations are the most common genetic alterations in breast cancer (Banerji et al., 2012; COSMIC). Their frequency in Neocent breast cancer tissues were 34.6%-38.5% and 19.2%-46.2% respectively and are present in approximately 69.2% of all Neocent breast cancers (n=26). These frequencies are slightly higher than the previous finding, which is 27%-37% for *PIK3CA* mutation and 27%-36% for *TP53* mutation, and they are present in approximately 55% to 65% of all breast cancers (Cancer Genome Atlas Network, 2012). However, some studies reported that *PIK3CA* mutation frequency in breast cancer is wide range from 8-40% (Samuels et al., 2004; Levine et al., 2005). 19.2% of Neocent breast cancer tissue has both *PIK3CA* and *TP53* mutations and 11.5% harboured two different *TP53* mutations.

The gain of *PIK3CA* gene copy number was approximately 7.7 %. Also, *PIK3CA* deletion was 7.7%. The *PIK3CA* CNA was not frequent as *PIK3CA* mutation and that supported by the finding of another study, which reported that the *PIK3CA* somatic mutation rather than a *PIK3CA* gene copy number gain is the frequent genetic alteration that contributes to the progression of breast cancer (Wu et al., 2005).

Several studies indicate the importance of *PIK3CA* mutation in determining the survival and treatment efficacy (Berns et al., 2007; Jensen et al., 2012). Some studies suggest the association of *PIK3CA* mutation with poor survival despite adjuvant chemotherapy and trastuzumab (Jensen et al., 2012). Some similar studies (Berns et al., 2007; Jensen et al., 2012; Cizkova et al., 2013) but not all (Esteva et al., 2010; Loi et al., 2013) have emphasised the prognostic value of *PIK3CA* mutation. Thesis results demonstrate the association of *PIK3CA* mutations with treatment response. Most of Neocent patients with *PIK3CA* mutation (70%) have significant radiological response regardless of treatment received. That could strength the studies, which suggest the association of *PIK3CA* mutation with better survival compared with *PIK3CA* wild-type breast cancers (Cizkova et al., 2012). Moreover, the presence of *PIK3CA* mutation in Neocent tissues associated with more favourable response to chemotherapy (75%) than endocrine therapy (66.7%). On the other hand, the presence of *TP53* mutations associated with endocrine treatment resistance, as only 42.9% of Neocent patients with *TP53* mutation and received endocrine therapy have a significant radiological response. That supported

by the finding of other study which have been reported the association between the presence of a *TP53* mutation and poor prognosis in patients treated with hormonal therapy (Kim et al., 2010). Also, another study where they found tumours with a *TP53* mutation have a more aggressive phenotype, and high histological grade, compared with those without a *TP53* mutation (Olivier et al., 2006). Moreover, the presence of a combination of *PIK3CA* and *TP53* mutations or two different *TP53* mutations in the same Neocent breast cancer tissue correlate with treatment resistance to either chemotherapy or endocrine therapy which indicated by no radiological response. Many studies investigate the correlation between the *TP53* status and prognosis. Most of these studies suggest that a *TP53* mutation is associated with poor prognosis (Pharoah et al., 1999), although contradictory results have also been reported, in particular for patients treated with chemotherapy (Varna et al., 2011). The reason behind that could be due to the effect of chemo-sensitivity of breast tumours by the presence of a *TP53* mutation (Bertheau et al., 2007). However, in our thesis the existence of the *TP53* mutation in Neocent tissues appeared to have no effect on chemotherapy response.

All the Neocent tissues with *PIK3CA* CNAs associated with significant radiological response regardless of treatment received. While, in the previous study, the *PIK3CA* CNV (gain) associated with resistance to PI3K inhibitors in breast cancer (Huw et al., 2013).

Based on ddPCR results, the *PIK3CA* mutation and *TP53* mutation fraction reduced after neoadjuvant chemotherapy and associated with radiological response and that could be supported by the study which suggests a loss of *PIK3CA* and *TP53* mutations after neoadjuvant chemotherapy are associated with better prognosis (Jiang et al., 2014). However, the frequency of *PIK3CA* mutations didn't change by neoadjuvant endocrine treatment and associated with no radiological response. On the other hand, the *TP53* mutation reduced after neoadjuvant endocrine therapy but has no effect on the radiological response.

The presence of *PIK3CA* or *TP53* mutations in cf-DNA was identified in previous studies. The breast cancer was shown to harbour the identical *PIK3CA* mutation identified in peripheral blood (Jelovac et al., 2014). Also, the *PIK3CA* mutation was detected in the cf-DNA of early breast cancer patients (Beaver et al., 2014). However, the clinicopathological features and prognostic significance of breast tumours with

PIK3CA mutations in the cf-DNA have not been documented (Beaver et al., 2014; Bettegowda et al., 2014). The *PIK3CA* and *TP53* mutation were detected in cf-DNA of metastatic breast cancer using NGS (Nakauchi et al., 2016). In the thesis pilot study, *PIK3CA* mutation was identified in 50% of cancer tissue DNA but only detected in 20% of the matched cf-DNA to the tumours harbour *PIK3CA* mutations, using qPCR with PNA. The lower frequency of *PIK3CA* mutation identified in cf-DNA using qPCR with PNA compared to other studies which reported *PIK3CA* mutation in cf-DNA in approximately 80 % of metastatic breast cancer patients whose tumours harbour *PIK3CA* mutations (Dawson et al., 2013b), could be due to either the fragmentation of cf-DNA or the presence of very small concentration of *PIK3CA* mutation in cf-DNA of early breast cancer compared to metastatic stage, which couldn't detected using qPCR with PNA. Moreover, whole-genome amplification (WGA) of cf-DNA may have biased this further. The *PIK3CA* mutation was detected in only 50% of metastatic cf-DNA harbour *PIK3CA* (c.3140A>G) confirmed by NGS, and none in the previously sequenced metastatic DNA with *PIK3CA* (c.1633G>A) and that could be again due to sample quality and fragmentation of cf-DNA. Because, the designed *PIK3CA* mutation assays was used in a recent study and identified *PIK3CA* (c.3140A>G) mutation in cf-DNA using qPCR with PNA and *PIK3CA* (c.1633G>A) mutation in cf-DNA using ddPCR (Guttery et al., 2015).

The *PIK3CA* and *TP53* genes are frequent cancer driven genes found in the whole exome sequencing analysis of ten Neocent breast cancer tissues, and that was supported by the finding of another research group (Nik-Zainal et al., 2016). Also, the whole exome sequencing of 10 Neocent tissues DNA confirmed the presence of *PIK3CA* mutation in the Neocent tissue of the same patient, which detects *PIK3CA* mutation in cancer tissue DNA samples with qPCR and ddPCR. However, it didn't detect *TP53* hotspot mutations in ten Neocent tissues DNA, which was detected with HRMC or ddPCR. Also, it didn't confirm the presence of *PIK3CA* (c.3140A>G) in one of Neocent tissue, which was detected with ddPCR in very low concentration.

Chapter 5. Conclusions and future direction.

5.1 Conclusions

The identification of specific and non-invasive biomarkers is needed to help in screening, diagnosis and management of breast cancer. Two circulating biomarkers currently under active investigation are circulating microRNAs (miRNAs) and circulating cell-free tumour DNA (cf-DNA). The aim was to profile circulating miRNAs in plasma of breast cancer and compare to healthy controls to determine whether there were any miRNAs that were differentially expressed in cancer. A second aim was to compare different methods to detect *PIK3CA* and *TP53* mutations in breast cancer tissues DNA and cf-DNA.

Many studies have suggested a role for circulating miRNAs as novel cancer biomarkers (Roth et al., 2010; Zhao et al., 2010 (a); Turchinovich et al., 2011). By analysing the expression levels of 384 miRNAs in breast cancer plasma pooled samples and compared to controls plasma pooled samples using TaqMan MicroRNA Array, there was expression difference of 49 miRNAs in plasma of breast cancer patients compared to female healthy controls ($P<0.05$). However, with further analysis of 11 selected miRNAs, there was an overlap of miRNA expression level between cancer patients and controls. Also, there was a wide range of variability in the expression level of the same miRNA in different control samples. The variability of miRNA expression in control plasma samples could suggest the biological difference of the expression level of particular miRNAs in different healthy controls, processing differences and/or white blood cell lysis (Kirschner et al., 2013). MiRNAs are not stable as previously known as the timing between blood collections and processing could affect the expression profile of miRNAs in plasma but not miRNAs extracted from exosomes. Also, the origin of plasma miRNA could be from blood cells rather than cancer miRNA, so some study provide specific practical recommendations to help minimise the variation attributable to plasma processing and platelet contamination (Cheng et al., 2013). These combined effects of several variables made results poorly reproducible and their translation into clinically useful applications not feasible (Cheng et al., 2013; Sourvinou et al., 2013; Mestdagh et al., 2014). It became evident that several variables (sample collection and storage, the differences in sample type (plasma or serum), sample processing protocols, RNA purification methods, quantification and normalisation can affect the final results and produce non-comparable or difficult-to-compare results (Jarry et al., 2014).

Therefore, standardisation of processes is required to minimise variation in liquid biopsies, particularly for miRNA (Page et al., 2013) and to achieve robust, accurate and reproducible miRNA profiles from collected samples in a simple manner required the development of a high-quality protocol to isolate and characterize miRNA from exosomes.

The second aim, which was to compare different methods to detect *PIK3CA* and *TP53* mutations in breast cancer tissues DNA and cf-DNA. The ddPCR method was found to have higher sensitivity in detecting low level *PIK3CA* (0.1%) or *TP53* (0.01%) mutations in breast cancer tissue than that achieved with qPCR or HRMC with PNA and DNA Sanger sequencing. Therefore, the ddPCR technology could become a very efficient and non-invasive tool in oncology, complementary to conventional diagnostic techniques. The qPCR with PNA and ddPCR assays showed similar sensitivity in detecting *PIK3CA* mutation in cell line DNA. However, the ddPCR has advantages over qPCR with PNA as it discovered additional mutations in breast cancer tissues not found by qPCR or Sanger sequencing and it can provide an accurate quantitative measurement of mutation. The qPCR with PNA is a sensitive method for detecting small mutant proportion but may be affected by DNA quality more than ddPCR as ddPCR based on analysis of each droplet generated. Therefore, the sensitivity of qPCR with PNA in detecting low-level *PIK3CA* was lower in cf-DNA compared to Neocent breast cancer tissue.

The frequency of *PIK3CA* mutation in breast cancer tissue was slightly higher with ddPCR (38.5%) compared to qPCR with PNA (34.6%). While the frequency of *TP53* mutation was greater with HRMC (46.2%) compared to ddPCR (19.2%). However, there was a wide variation and failure in HRMC results which may cause false positive results which weren't identified by ddPCR. Therefore, identifying *TP53* mutation using HRMC with PNA needs further validation. The frequency of *PIK3CA* somatic mutation was more frequent than the gain (7.7%) or deletion (7.7%) of gene copy number. The presence of *PIK3CA* mutation or gain or deletion of *PIK3CA* copy number may correlate with the response to either chemotherapy or endocrine therapy. While, the presence of the *TP53* mutation may correlate with endocrine resistance. Moreover, the presence of a combination of *PIK3CA* and *TP53* mutations or two different mutations in

the same breast cancer tissue correlate with treatment resistance to either chemotherapy or endocrine therapy which indicated by no radiological response.

In conclusion, the results in this thesis have shown that microRNAs are challenging to investigate through liquid biopsy, whereas cf-DNA analysis seems more robust and encouraging and ddPCR method could become a very efficient and non-invasive tool in oncology.

5.2 Future direction

5.2.1 MiRNAs

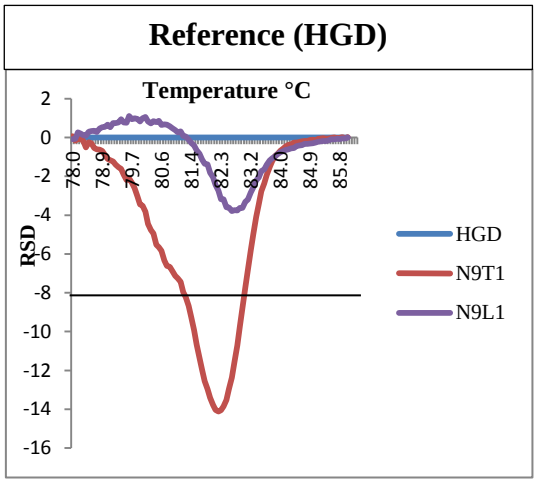
In order to take this study forward, it would be important to validate the exosome extraction method to extract miRNAs and repeat the experiment in a larger samples number with more controlling during processing of samples to investigate the reasons for the variation in control samples. After this the functional significance of specific miRNA changes could be studied in cell line models through gene knockdown using RNA interference (RNAi) experiments.

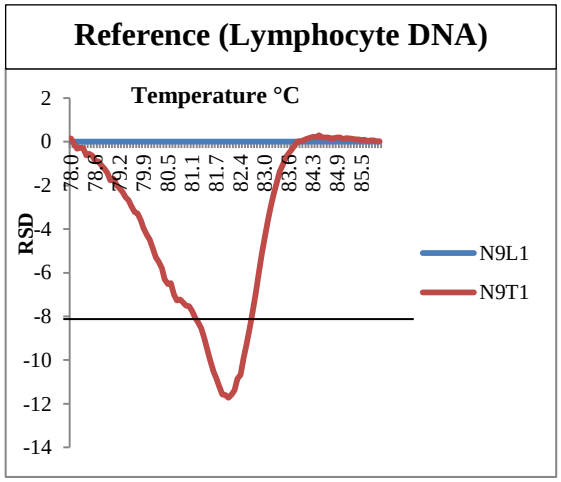
5.2.2 *PIK3CA* and *TP53* somatic mutations

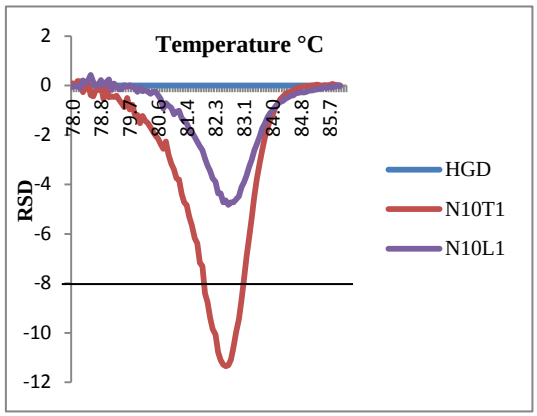
Regarding technical approaches, further validation of HRMC method is desired for the screening of *TP53* mutations. Future studies can also address whether *PIK3CA* and *TP53* mutations can be detected in the paired pre-amplified and non-amplified cf-DNA using ddPCR assays to identify ddPCR sensitivity in detecting these mutations in cf-DNA and their frequency as they may serve as biomarkers to detect early stage breast cancer. After that, it would be important to confirm the presence of *PIK3CA* or *TP53* mutations in cf-DNA with NGS and compare all the results with whole exome sequence results, followed by validation of the results with a larger number of patients with early breast cancer and healthy female controls. Statistical analysis of all the collected data would be important and assess the association of circulating biomarkers with the diagnostic histology of tissue biopsy and/or mammography.

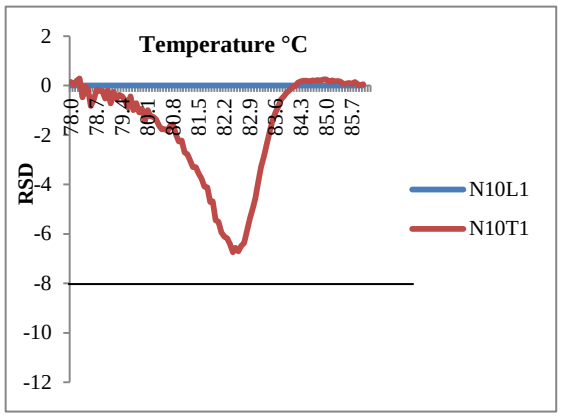
Appendices

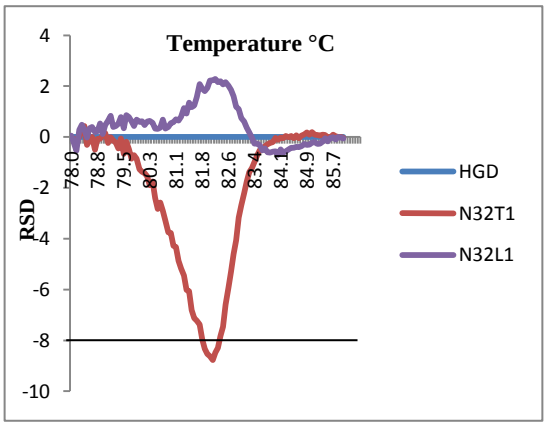
Appendix 1:

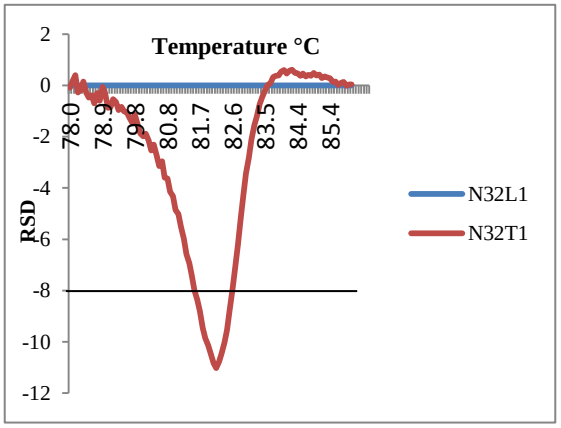

Reference (HGD)
Temperature °C
RSD
HGD
N9T1
N9L1

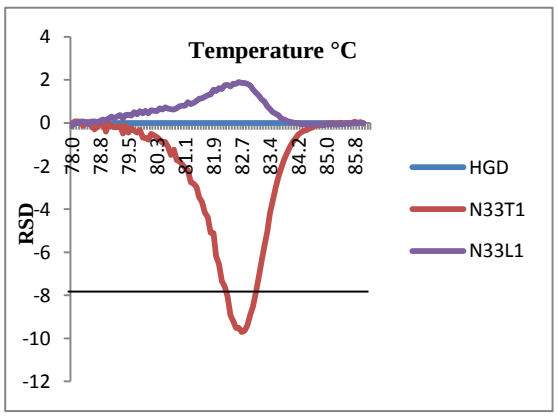

Reference (Lymphocyte DNA)
Temperature °C
RSD
N9L1
N9T1

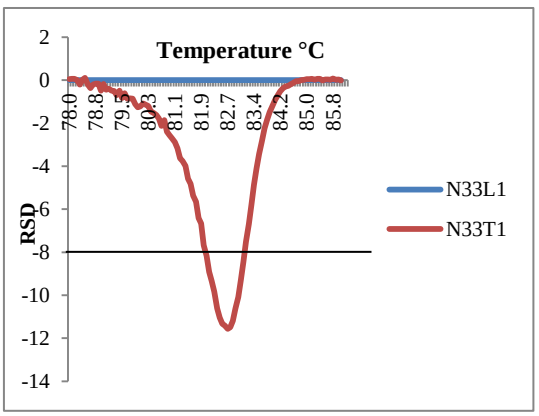

Temperature °C
RSD
HGD
N10T1
N10L1

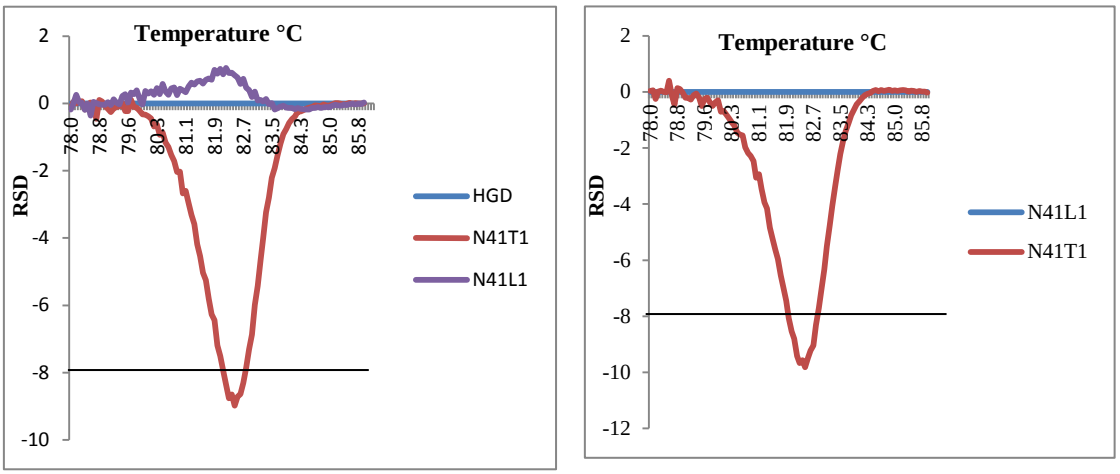

Temperature °C
RSD
N10L1
N10T1


Temperature °C
RSD
HGD
N32T1
N32L1


Temperature °C
RSD
N32L1
N32T1

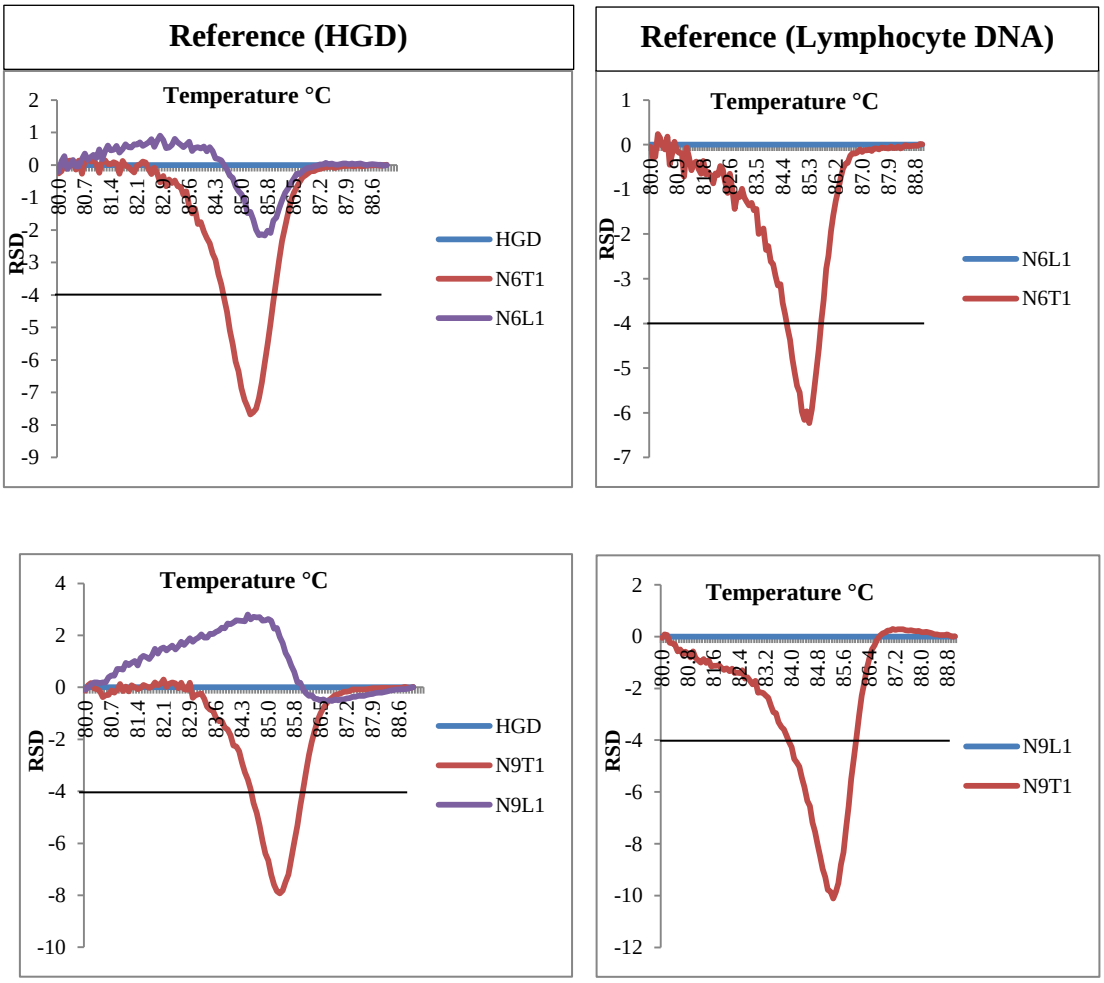

Temperature °C
RSD
HGD
N33T1
N33L1

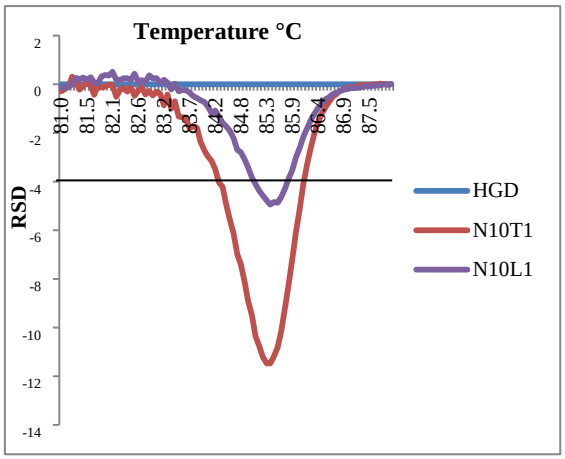

Temperature °C
RSD
N33L1
N33T1

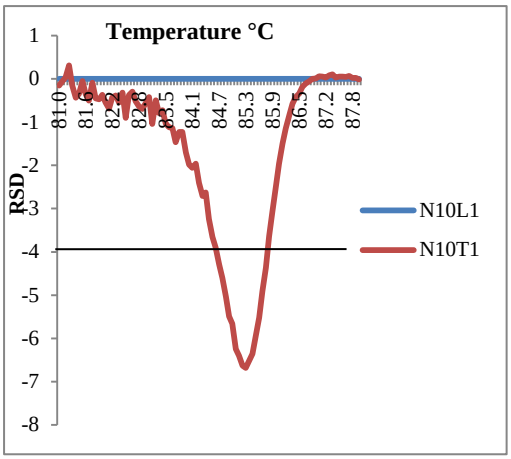


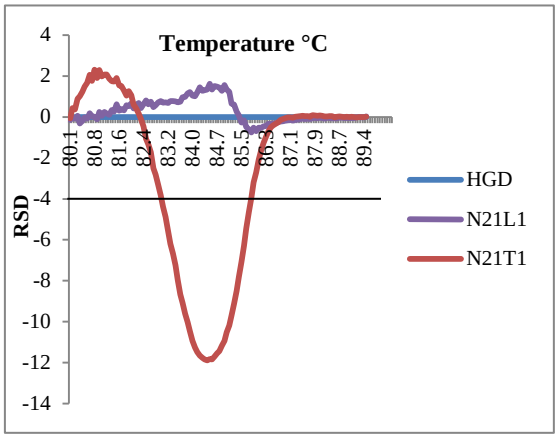
Appendix 1. Neocent tissue's DNAs with *TP53* (c.818G>A) mutation. Relative fluorescence signal differences between Neocent tissues DNAs and HGD or matched lymphocytes DNAs. The threshold line is at -8 relative signal difference for *TP53* (c.818G>A) mutation. RSD: relative signal difference.

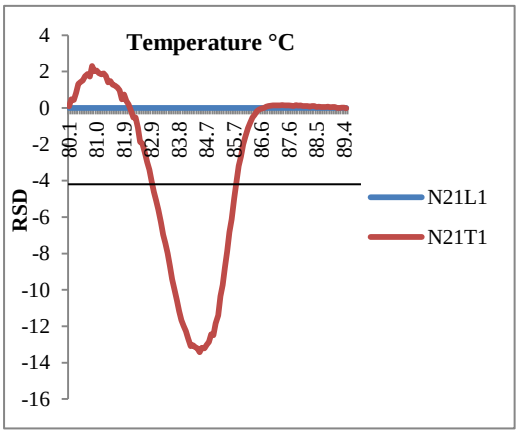
Appendix 2:

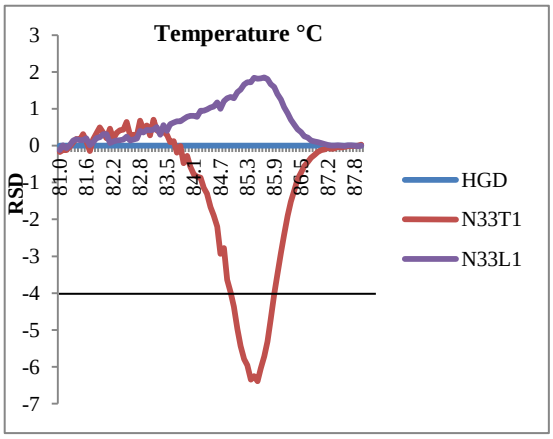


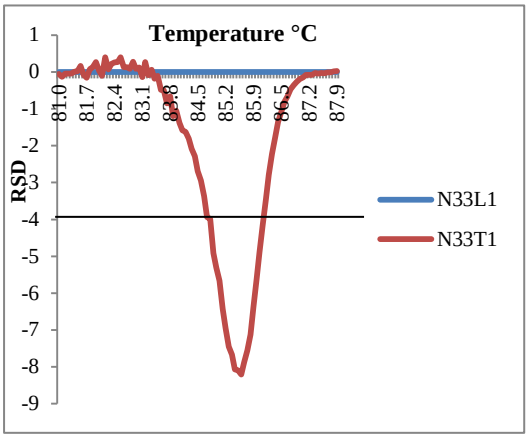

Temperature °C
RSD
HGD
N10T1
N10L1

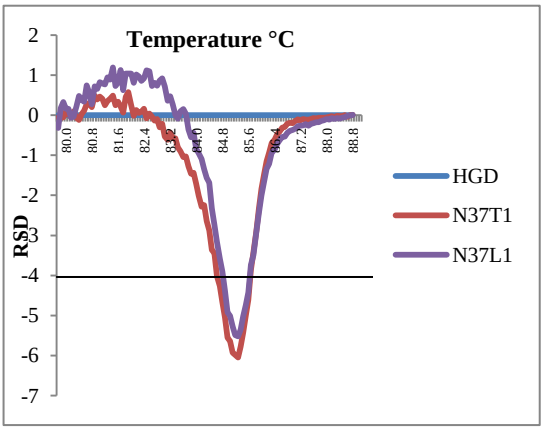

Temperature °C
RSD
N10L1
N10T1

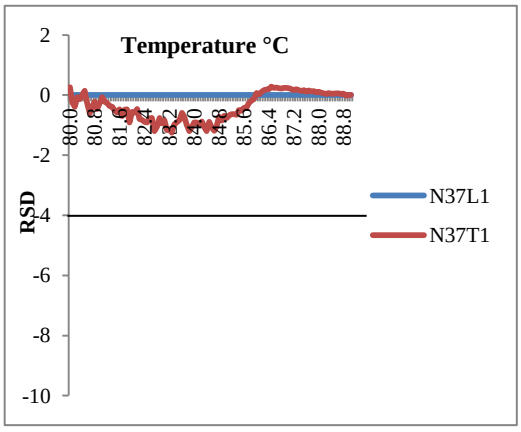

Temperature °C
RSD
HGD
N21L1
N21T1

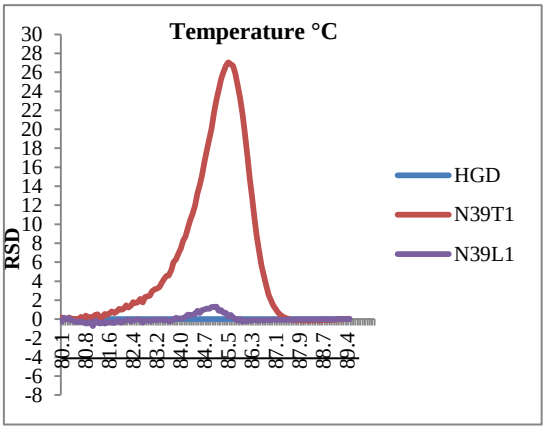

Temperature °C
RSD
N21L1
N21T1

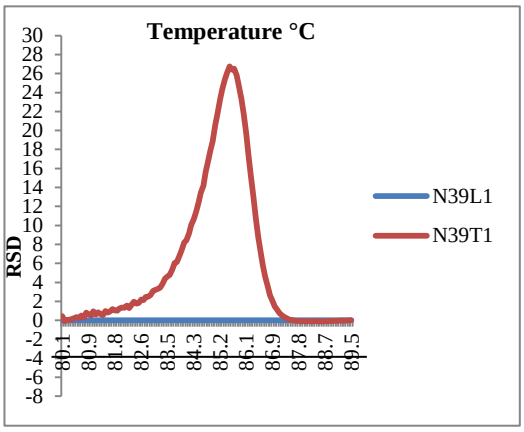

Temperature °C
RSD
HGD
N33T1
N33L1

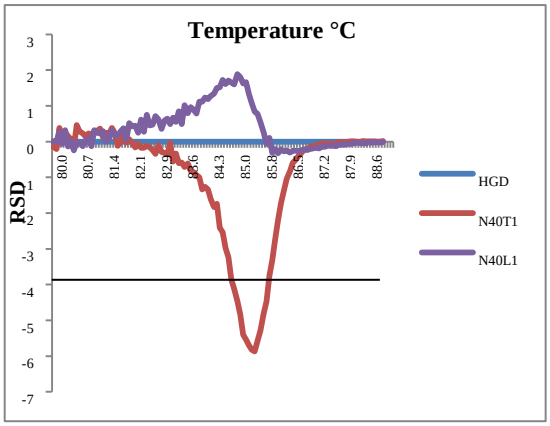

Temperature °C
RSD
N33L1
N33T1

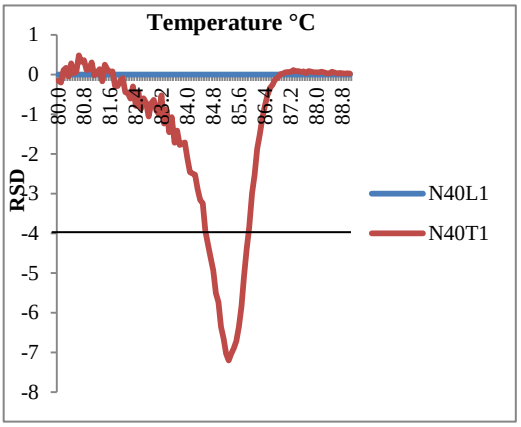

Temperature °C
RSD
HGD
N37T1
N37L1

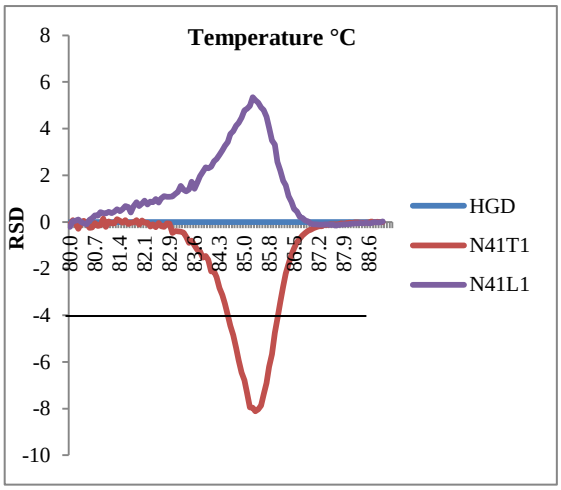

Temperature °C
RSD
N37L1
N37T1

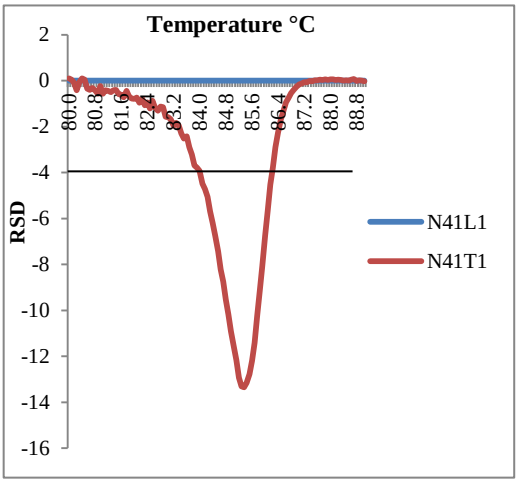

Temperature °C
RSD
HGD
N39T1
N39L1

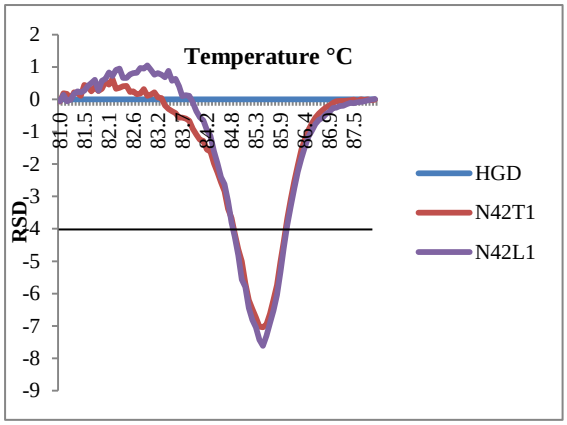

Temperature °C
RSD
N39L1
N39T1

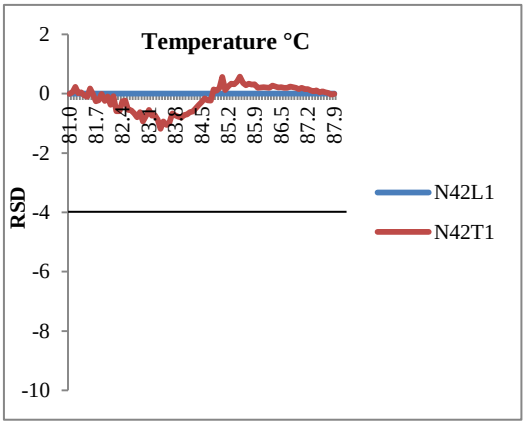

Temperature °C
RSD
HGD
N40T1
N40L1


Temperature °C
RSD
N40L1
N40T1


Temperature °C
RSD
HGD
N41T1
N41L1


Temperature °C
RSD
N41L1
N41T1


Temperature °C
RSD
HGD
N42T1
N42L1


Temperature °C
RSD
N42L1
N42T1

Appendix 2. Neocent tissue's DNAs with *TP53* (c.524G>A) mutation. Relative fluorescence signal differences between Neocent tissues DNAs and HGD or matched lymphocytes DNAs. The threshold line is at -4 relative signal difference for *TP53* (c.524G>A) mutation. RSD; relative signal difference.

Appendix 3: Ethical approval letter of BSMS study

NHS
Health Research Authority

NRES Committee London - Riverside
Bristol Research Ethics Committee Centre
Level 3 Block B
Whitefriars
Lewins Mead
Bristol
BS1 2NT

Telephone: 0117 342 1385
Facsimile: 0117 342 0445

24 January 2013

Professor Justin Stebbing
Professor of Cancer Medicine and Medical Oncology
Imperial Healthcare NHS Trust
Dept of Medical Oncology, 1st Floor, East Wing
Charing Cross Hospital,
Fulham Palace Road, London
W6 8RF

Dear Professor Stebbing

Study title: **Breast Screening and Monitoring Study**
REC reference: **12/LO/2019**
IRAS project ID: **118701**

Thank you for your letter of 21 January 2013, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

We plan to publish your research summary wording for the above study on the NRES website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the Co-ordinator Miss Tina Cavaliere, nrescommittee.london-riverside@nhs.net.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS sites

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Covering Letter		03 December 2012
Covering Letter		20 January 2013
Investigator CV		03 December 2012
Letter from Sponsor		27 November 2012
Letter of invitation to participant	1	06 November 2012
Other: Letter confirming review of protocol (Brighton and Sussex Medical School)		28 November 2012
Other: Letter from Royal United Hospital		27 November 2012
Other: Letter of support from University of Southampton		18 November 2012

Participant Consent Form: Healthy Volunteers	2 - Clean & Tracked	21 January 2013
Participant Consent Form: Patients	2 - Clean & Tracked	21 January 2013
Participant Information Sheet: Healthy Volunteers	2 - Clean & Tracked	21 January 2013
Participant Information Sheet: Patients	2 - Clean & Tracked	21 January 2013
Protocol	1	06 November 2012
REC application		03 December 2012
Referees or other scientific critique report		18 November 2012
Response to Request for Further Information		21 January 2013

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

12/LO/2019 **Please quote this number on all correspondence**

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

A Research Ethics Committee established by the Health Research Authority

Yours sincerely

pp.

Dr Sabita Uthaya
Chair

Email:nrescommittee.london-riverside@nhs.net

Enclosures: "After ethical review – guidance for researchers"

Copy to: *Miss Becky Ward, Imperial College London and Imperial College Healthcare NHS Trust*

References

Ahmed, D., Eide, P.W., Eilertsen, I.A., Danielsen, S.A., Eknaes, M,, Hektoen, M.Lind, G.E. and Lothe, R.A., 2013. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis, **2**, pp. e71.

Ambros,V., 2003. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell*, **113**, pp. 673–676.

Anker, P., Mulcahy, H., Chen, X.Q. and Stroun, M., 1999. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. *Cancer Metastasis Rev*,**18**, pp. 65–73.

Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., et al., 2011. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A*, **108**, pp. 5003–5008.

Asaga, S., Kuo, C., Nguyen, T., Terpenning, M., Giuliano, A.E. and Hoon, DS., 2011. Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. *Clin Chem*, **57**(1), pp. 84-91.

Ayub, S.G., Rasool, S., Ayub, T., Khan, S.N., Wani, K.A. and Andrabi, K.I., 2014. Mutational analysis of the BRCA2 gene in breast carcinoma patients of Kashmiri descent. *Molecular medicine reports*, **9**(2), pp. 749-53.

Bader, A.G., Kang, S., Zhao, L., et al., 2005. Oncogenic PI3K deregulates transcription and translation. *Nat Rev Cancer*, **5**, pp. 921–929.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., et al., 2008. The impact of microRNAs on protein output. *Nature*, **455**, pp. 64–71.

Bai, X., Zhang, E., Ye, H., Nandakumar, V., Wang, Z., Chen, L., et al., 2014. PIK3CA and TP53 gene mutations in human breast cancer tumors frequently detected by ion torrent DNA sequencing. *PLoS One*, **9**, pp. e99306.

Baker, L., Quinlan, P.R., Patten, N., Ashfield, A., Birse-Stewart-Bell, L.J., McCowan, C. et al., 2010. p53 mutation, deprivation and poor prognosis in primary breast cancer. *Br J Cancer*, **102**, pp. 719–26.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., et al., 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, **486**, pp. 405–9.

Barbareschi, M., Buttitta, F., Felicioni, L., Cotrupi, S., Barassi, F., Del, G.M., et al, 2007. Different prognostic roles of mutations in the helical and kinase domains of the PIK3CA gene in breast carcinomas. *Clin Cancer Res*, **13**, pp. 6064–9.

Barnes, T., Bell, K., DiSebastiano, K.M., Vance, V., Hanning, R., Russell, C., Dubin, J.A., Bahl, M., Califaretti, N., Campbell, C.. and Mourtzakis, M. 2014. Plasma amino acid profiles of breast cancer patients early in the trajectory of the disease differ from healthy comparison groups. *Appl Physiol Nutr Metab*, **39**(6), pp. 740–744.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, pp. 281–97.

Beaver, J.A., Jelovac, D., Balukrishna, S., Cochran, R.L., Croessmann, S., et al., 2014. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res*, **20**, pp. 2643–2650.

Berg, W.A., Gutierrez, L., NessAiver, M.S., Carter, W.B., Bhargavan, M., Lewis, R.S., Ioffe, O.B., 2004. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology*, **233**(3), pp. 830–849.

Berns, K., Horlings, H.M., Hennessy, B.T., Madiredjo, M., Hijmans, E.M., Beelen, K., et al., 2007. A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell*, **12**, pp. 395–402.

Bertheau, P., Turpin, E., Rickman, D.S., Espie, M., de Reynies, A., Feugeas, J.P., Plassa, L.F., Soliman, H., Varna, M., de Roquancourt, A., Lehmann-Che, J., Beuzard, Y., Marty, M., Misset, J.L., Janin, A. and de The, H., 2007. Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin–cyclophosphamide regimen. *PLoS Med*, **4** (3), pp. e90.

Bertheau, P., Espie, M., Turpin, E., Lehmann, J., Plassa, L.F., Varna, M., Janin, A. and de The, H., 2008. TP53 status and response to chemotherapy in breast cancer. *Pathobiology*, 75 (2), pp. 132–139.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B, Alani, R.M., Antonarakis, E.S., Azad, N.S., Bardelli, A., Brem, H., Cameron, J.L. Lee, C.C., Fecher, L.A., Gallia, G.L., Gibbs, P., Le, D., Giuntoli, R.L., Goggins, M., Hogarty, M.D., Holdhoff, M., Hong, S.M., Jiao, Y., Juhl, H.H., Kim, J.J., Siravegna, G., Laheru, D.A., Lauricella, C., Lim, M., Lipson, E.J., Marie, S.K., Netto, G.J., Oliner, K.S., Olivi, A., Olsson, L., Riggins, G.J., Sartore-Bianchi, A., Schmidt, K., Shih, I.M., Oba-Shinjo, S.M., Siena, S., Theodorescu, D., Tie, J., Harkins, T.T., Veronese, S., Wang, T.L., Weingart, J.D., Wolfgang, C.L., Wood, L.D., Xing, D., Hruban, R.H., Wu, J., Allen, P.J. Schmidt, C.M., Choti, M.A.,

Velculescu, V.E., Kinzler, K.W., Vogelstein, M.B., Papadopoulos, N. and Diaz, L.A., 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*, **6**(224), pp. 224ra224.

Black, J.S., Salto-Tellez, M., Mills, K.I. and Catherwood, M.A., 2015. The impact of next generation sequencing technologies on haematological research. *Pathogenesis*, **2**, pp. 9-16.

Borresen-Dale, A.L. 2003. TP53 and breast cancer. *Hum Mutat*, **21** (3), pp. 292–300.

Bossi, G., Lapi, E., Strano, S., Rinaldo, C., Blandino, G., and Sacchi, A. 2006. Mutant p53 gain of function: reduction of tumor malignancy of human cancer cell lines through abrogation of mutant p53 expression. *Oncogene*, **25**, pp. 304–309.

Bracken, C.P., Khew-Goodall, Y. and Goodall, G.J., 2015. Network-Based Approaches to Understand the Roles of miR-200 and other microRNAs in Cancer. *Cancer Res*, **75**, pp. 2594-2599.

Brase, J.C., Wuttig, D., Kuner, R. and Sultmann, H., 2010. Serum microRNAs as non-invasive biomarkers for cancer. *Mol Cancer*, **9**, pp. 306 .

Britton, P., Warwick, J., Wallis, M.G., O'Keeffe, S,. Taylor, K., et al., 2012. Measuring the accuracy of diagnostic imaging in symptomatic breast patients: team and individual performance. *Br J Radiol*, **85**(1012), pp. 415-22.

Cain, R.J. and Ridley, A.J., 2009. Phosphoinositide 3-kinases in cell migration. *Biol Cell*, **101**(1), pp. 13–29.

Calderon-Garciduenas, A.L., Ruiz-Flores, P., Cerda-Flores, RM. and Barrera-Saldana, HA., 2005. Clinical follow up of mexican women with early onset of breast cancer and mutations in the BRCA1 and BRCA2 genes. *Salud publica de Mexico*, **47**(2), pp. 110-5.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., et al., 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*, **101**, pp. 2999–3004.

Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., et al., 2005. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*, **353**, pp. 1793–1801.

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

Casilli, F., Tournier, I., Sinilnikova, O.M., Coulet, F., Soubrier, F., Houdayer, C., Hardouin, A., Berthet, P., Sobol, H., Bourdon, V., Muller, D., Fricker, J.P., Capoulade-Metay, C., Chompret, A., Nogues, C., Mazoyer, S., Chappuis, P., Maillet, P., Philippe, C., Lortholary, A., Gesta, P., Bézieau, S., Toulas, C., Gladieff, L., Maugard, C.M., Provencher, D.M., Dugast, C., Nguyen, T.D., Faivre, L., et al., 2006. The contribution of germline rearrangements to the spectrum of *BRCA2* mutations. *J Med Genet*, **43**, pp. e49.

Chan, J.A., Krichevsky, A.M., and Kosik, K.S. 2005. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res*, **65**, pp. 6029–6033.

Chang, L., Shrestha, S., LaChaud, G., Scott, M.A. and James, A.W., 2015. Review of microRNA in osteosarcoma and chondrosarcoma. *Med Oncol*, 32, pp. 613.

Checka, C.M., Chun, J.E., Schnabel, F.R., et al., 2012. The relationship of mammographic density and age: implications for breast cancer screening. *AJR Am J Roentgenol*, **198**, pp. W292–5.

Chen, X. et al., 2008. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research*, **18**, pp. 997–1006.

Chen, W.X., Hu, Q., Qiu, M.T., Zhong, S.L., Xu, J.J. et al., 2013. miR-221/222: promising biomarkers for breast cancer. *Tumour Biol*, **34**, pp. 1361–1370.

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

Cheng, J., and Haas, M., 1990. Frequent mutations in the p53 tumor suppressor gene in human leukemia T-cell lines. *Mol Cell Biol*,**10**, pp. 5502-9.

Cheng, K.C. and Loeb, L.A., 1997. Genomic stability and instability: a working paradigm. *Curr. Top. Microbiol. Immunol.*, **221**, pp. 5–18.

Cheng, H.H., Yi, H.S., Kim, Y., Kroh, E.M., Chien, J.W., Eaton, K.D., et al., 2013. Plasma processing conditions substantially influence circulating microRNA biomarker levels. *PLoS One*, **8**, pp. e64795.

Chin, L.J. and Slack, F.J. 2008. A truth serum for cancer—microRNAs have major potential as cancer biomarkers. *Cell Res*, **18**, pp. 983–984.

Chiu, T.W., Young, R., Chan, L.Y., Burd, A. and Lo, D.Y., 2006. Plasma cell-free DNA as an indicator of severity of injury in burn patients. *Clin Chem Lab Med*, **44**, pp. 13-17.

Cho, Y., Gorina, S., Jeffrey, P.D. and Pavletich, N.P., 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*, **265**, pp. 346-55.

Cizkova, M., Susini, A., Vacher, S., et al., 2012. PIK3CA mutation impact on survival in breast cancer patients and in ERalpha, PR and ERBB2-based subgroups. *Breast Cancer Res*, **14**(1), pp. R28.

Cizkova, M., Dujaric, M.E., Lehmann-Che, J., Scott, V., Tembo, O., et al., 2013. Outcome impact of PIK3CA mutations in HER2-positive breast cancer patients treated with trastuzumab. *Br J Cancer*, **108**, pp. 1807–1809.

Cleator, S.J., Powles, T.J., Dexter, T., Fulford, L., Mackay, A., Smith, I.E., Valgeirsson, H., Ashworth, A. and Dowsett, M., 2006. The effect of the stromal component of breast tumours on prediction of clinical outcome using gene expression microarray analysis. *Breast Cancer Res*, **8** (3), pp. R32.

Cookson, V.J., Bentley, M.A., Hogan, B.V., Horgan, K., Hayward, B.E., Hazelwood, L.D. and Hughes, T.A., 2012. Circulating microRNA profiles reflect the presence of breast tumours but not the profiles of microRNAs within the tumours. *Cell Oncol (Dordr)*, **35**(4), pp. 301-8.

Corsetti, V., Houssami, N., Ghirardi, M., et al., 2011. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *Eur J Cancer*, **47**, pp. 1021-1026.

Cortez, M.A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A.K., et al., 2011. MicroRNAs in body fluids-the mix of hormones and biomarkers. *Nat Rev Clin Oncol*, **8**, pp. 467–477.

COSMIC (www.cancer.sanger.ac.uk/COSMIC).

Creemers, E.E., Tijsen, A.J. and Pinto, Y.M., 2012. Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease. *Circulation research*, **110**, pp. 483–495.

Croce, C.M., 2009. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet*, **10**, pp. 704–714.

Crowder, R.J., Phommaly, C., Tao, Y., Hoog, J., Luo, J., Perou, C.M., et al., 2009. PIK3CA and PIK3CB inhibition produce synthetic lethality when combined with estrogen deprivation in estrogen receptor-positive breast cancer. *Cancer Res*, **69**, pp. 3955e62.

Crowley, E., Di Nicolantonio, F., Loupakis, F., Bardelli, A., 2013. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*, **10**(8), pp. 472–484.

Cuk, K., Zucknick, M., Heil, J., Madhavan, D., Schott, S., Turchinovich, A., et al., 2013. Circulating microRNAs in plasma as early detection markers for breast cancer. *International Journal of Cancer*, **132**(7), pp. 1602-12.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S.; METABRIC Group, Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A,. Borresen-Dale, A.L., Brenton, J.D., Tavare, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**(7403), pp. 346-52.

Dawson, S.J., Rueda, O.M., Aparicio, S. and Caldas, C., 2013a. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*, **32**(5), pp. 617-28. To remove

Dawson, S.J., Tsui, D.W., Murtaza, M., Biggs, H., Rueda, O.M., Chin, S.F., et al., 2013b. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med*, **368**, pp. 1199–209.

DeLeo, A.B., Jay, G., Appella, E., Dubois, G.C., Law, LW. and Old, L.J., 1979. Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proc Natl Acad Sci*, **76**, pp. 2420–2424.

Deligezer, U., Eralp, Y., Akisik, E.Z., Akisik, E.E., Saip, P. and Topuz, E., 2008. Effect of adjuvant chemotherapy on integrity of free serum DNA in patients with breast cancer. *Ann N Y Acad Sci*, **1137**, pp. 175–9.

Diaz, L.A. and Bardelli, A., 2014. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol*, **32**(6), pp. 579–586.

Diehl, F., Schmidt, K., Choti, M.A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S.A., Kinzler, K.W., Vogelstein, B. and Diaz, L.A., 2008. Circulating mutant DNA to assess tumor dynamics. *Nat Med*, **14**(9), pp. 985–990.

Diskin, S.J., Hou, C., Glessner, J.T., Attiyeh, E.F., Laudenslager, M., Bosse, K., Cole, K., Mosse, Y.P., Wood, A., Lynch, J.E., Pecor, K., Diamond, M., Winter, C., Wang, K., Kim, C., Geiger, E.A., McGrady, P.W., Blakemore, A.I., London, W.B., Shaikh, T.H., Bradfield, J., Grant, S.F., Li, H., Devoto, M., Rappaport, E.R., Hakonarson, H. and Maris, J.M., 2009. Copy number variation at 1q21.1 associated with neuroblastoma. *Nature*, **459**, pp. 987-991.

Done, S.J., Arneson, C.R., Ozcelik, H., Redston, M. and Andrulis, I.L., 2001a. P53 protein accumulation in non-invasive lesions surrounding p53 mutation positive invasive breast cancers. *Breast Cancer Res. Treat*, **65** (2), pp. 111–118.

Done, S.J., Eskandarian, S., Bull, S., Redston, M. and Andrulis, I.L., 2001b. p53 missense mutations in microdissected high-grade ductal carcinoma in situ of the breast. *J Natl Cancer Inst*, **93**, pp. 700–4.

Dong, W., Yongjun, L., Nan, D., Junyun, W., Qiong, Y., Yaran, Y., Yanming, L., Xiangdong, F. and Hua, Z., 2015. Molecular networks and mechanisms of epithelial-mesenchymal transition regulated by miRNAs in the malignant melanoma cell line. *Yi Chuan*, **37**, pp. 673-682.

Duffy, M.J., 1999. CA 15–3 and related mucins as circulating markers in breast cancer. *Ann Clin Biochem*, **36** (Pt 5), pp. 579–586.

Dupont Jensen, J., Laenkholm, A.V., Knoop, A., et al., 2011. PIK3CA mutations may be discordant between primary and corresponding metastatic disease in breast cancer. *Clin Cancer Res*, **17**, pp. 667–77.

Dutta, K.K., Zhong, Y., Liu, Y.T., Yamada, T., Akatsuka, S., Hu, Q., Yoshihara, M., Ohara, H., Takehashi, M., Shinohara, T., Masutani, H., Onuki, J. and Toyokuni, S., 2007. Association of microRNA-34a overexpression with proliferation is cell type-dependent. *Cancer Sci*, **98**, pp. 1845-1852.

Duttagupta, R., Jiang, R., Gollub, J., Getts, R.C. and Jones, K.W., 2011. Impact of cellular miRNAs on circulating miRNA biomarker signatures. *PLoS One*, **6**, pp. e20769.

Edwards, J., Krishna, N.S., Witton, C.J., et al., 2003. Gene amplifications associated with the development of hormone-resistant prostate cancer. *Clin Cancer Res*, **9**, pp. 5271–5281.

Eikesdal, H.P., Knappskog, S., Aas, T., Lonning, P.E., 2014. TP53 status predicts long-term survival in locally advanced breast cancer after primary chemotherapy. *Acta Oncol*,**53**,pp.1347–55.

Engelman, J.A., Luo, J., Cantley, L.C., 2006. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat Rev Genet*, **7**, pp. 606–19.

Engelman, J.A., 2009. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nat Rev Cancer*, **9**(8), pp. 550–62.

Ensembl Genome Browser website *(www.ensembl.org).*

Esquela-Kerscher, A. and Slack, F.J., 2006. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*, **6**, pp. 259-269.

Esteva, F.J., Guo, H., Zhang, S., Santa-Maria, C., Stone, S., et al., 2010. PTEN, PIK3CA, p-AKT, and pp70S6K status: association with trastuzumab response and survival in patients with HER2-positive metastatic breast cancer. *Am J Pathol*, **177**, pp. 1647–1656.

Etheridge, A., Gomes, C.P., Pereira, R.W., Galas, D. and Wang, K., 2013. The complexity, function and applications of RNA in circulation. *Front Genet*, **4**, pp. 115

Fahy, B.N., Bold, R.J., Schneider, P.D., Khatri, V. and Goodnight, J.E., 2001. Costbenefit analysis of biopsy methods for suspicious mammographic lesions; discussion 994–5. *Arch Surg*, **136**(9), pp. 990-994.

Ferracin, M., Lupini, L., Salamon, I., Saccenti, E., Musa, G., Zagatti, B., et al., 2015. Absolute quantification of cell-free microRNAs in cancer patients. *Oncotarget*, **6**(16), pp. 14545-55.

Feuk, L., et al., 2006. Structural variation in the human genome. *Nature Reviews Genetics*, **7**, pp. 85–97.

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

Fleischhacker, M. and Schmidt, B., 2008. Cell-free DNA resuscitated for tumor testing. *Nature Medicine*, **14**, pp. 914 – 915.

Freed-Pastor, W.A. and Prives, C., 2012. Mutant p53: one name, many proteins. *Genes & development*, **26**(12), pp. 1268-86.

Fujimoto, A., O'Day, S.J., Taback, B., et al., 2004. Allelic imbalance on 12q22–23 in serum circulating DNA of melanoma patients predicts disease outcome. *Cancer Res*, **64**, pp. 40858.

Gao, J., Wu, H., Shi, X., Huo, Z., Zhang, J. and Liang, Z., 2016. Comparison of Next-Generation Sequencing, Quantitative PCR, and Sanger Sequencing for Mutation Profiling of EGFR, KRAS, PIK3CA and BRAF in Clinical Lung Tumors. *Clin Lab*, **62**(4), pp. 689-96.

Giacona, M.B., Ruben, G.C., Iczkowski, K.A., et al., 1998. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas*, **17**, pp. 89–97.

Goebel, G., Zitt, M. and Muller, H.M., 2005. Circulating nucleic acids in plasma or serum (CNAPS) as prognostic and predictive markers in patients with solid neoplasias. *Dis Markers*, **21**, pp. 105-120.

Gormally, E., Caboux, E., Vineis, P., et al., 2007. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutat Res*, **635**, pp. 105–17.

Greenblatt, M.S., Bennett, W.P., Hollstein, M. and Harris, C.C., 1994. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res*, **54**, pp. 4855–4878.

Gurtner, A., Starace, G., Norelli, G., Piaggio, G., Sacchi, A. and Bossi, G., 2010. Mutant p53-induced up-regulation of mitogen-activated protein kinase kinase 3 contributes to gain of function. *J Biol Chem*, **285**(19), pp. 14160-9.

Guttery, D.S., Page, K., Hills, A., Woodley, L., Marchese, S.D., Rghebi, B., Hastings, R.K., Luo, J., Pringle, J.H., Stebbing, J., Coombes, R.C., Ali, S. and Shaw, J.A., 2015. Noninvasive detection of activating estrogen receptor 1 (ESR1) mutations in estrogen receptor-positive metastatic breast cancer. *Clin Chem*, **61**(7), pp. 974-82.

Gymnopoulos, M., Elsliger, M.A. and Vogt, P.K., 2007. Rare cancerspecific mutations in PIK3CA show gain of function. *Proc Natl Acad Sci USA*, **104**(13), pp. 5569–5574.

Hainaut, P. and Hollstein, M., 2000. p53 and human cancer: the first ten thousand mutations. *Adv. Cancer Res*, **77**, pp. 81–137.

Hamam, D. et al., 2014. microRNA-320/RUNX2 axis regulates adipocytic differentiation of human mesenchymal (skeletal) stem cells. *Cell death & disease*, **5**, pp. e1499.

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell*, **144**, pp. 646–74.

Hannafon, B.N., Sebastiani, P., de Las Morenas, A., Lu, J. and Rosenberg, C.L., 2011. Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer. *Breast Cancer Res*, **13**(2), pp. R24.

Hao, W., Zhang, X., Xiu, B., Yang, X., Hu, S., Liu, Z., Duan, C., Jin, S., Ying, X., Zhao, Y., Han, X., Hao, X., Fan, Y., Johnson, H., Meng, D., Persson, J.L., Zhang, H., Feng, X. and Huang, Y., 2016. Vitronectin: a promising breast cancer serum biomarker for early diagnosis of breast cancer in patients. *Tumor Biol*, **37**(7), pp. 8909-16.

Harris, L., Fritsche, H., R., Norton, L., Ravdin, P., Taube, S., et al., 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*., **25**, pp. 5287–312.

He, L. and Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*, **5**, pp. 522–531.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W.,Hannon, G.J., et al., 2005. A microRNA polycistron as a potential human oncogene. *Nature*, **435**, pp. 828–833.

Helleday, T., Eshtad, S. and Nik-Zainal, S., 2014. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet*, **15**, pp. 585–598.

Heneghan, H.M., Miller, N., Lowery, A.J., Sweeney, K.J., Newell, J. and Kerin, M.J., 2010 a. Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Ann Surg*, **251**(3), pp. 499-505.

Heneghan, H.M., Miller, N., Kelly, R., Newell, J. and Kerin, M.J., 2010 b. Systemic miRNA-195 differentiates breast cancer from other malignancies and is a potential biomarker for detecting noninvasive and early stage disease. *Oncologist*, **15**(7), pp. 673-682.

Heneghan, H.M., Miller, N. and Kerin, M.J., 2011. Circulating microRNAs: promising breast cancer Biomarkers. *Breast Cancer Res*, **13**, pp. 402.

Henry, N.L. and Hayes, D.F., 2012. Cancer biomarkers. *Mol Oncol*, **6**, pp. 140–6.

Herrmann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T. and Voelkerding, K.V., 2006. Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clin Chem*, **52**, pp. 494e503

Hindson, B.J. Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., et al., 2011. High-throughput droplet digital pcr system for absolute quantitation of DNA copy number. *Analytical chemistry*, **83**, pp. 8604-10.

Ho, G.H., Calvano, J.E., Bisogna, M., Borgen, P.I., Rosen, P.P., Tan, L.K. and Van Zee, K.J., 2000. In microdissected ductal carcinoma in situ, HER-2/neu amplification, but not p53 mutation, is associated with high nuclear grade and comedo histology. *Cancer*, **89** (11), pp. 2153–2160.

Hollestelle, A., Elstrodt, F., Nagel, J.H.A., Kallemeijn, W.W., Schutte, M.l., 2007. Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol. Cancer Res*, **5**, pp.195e201.

Howlader, N., Noone, A., Krapcho, M., Garshell, J., Miller, D., Altekruse S., et al., 2014. SEER Cancer Statistics Review, 1975–2011. (http://seer.cancer.gov/csr).

Hu, Z., Dong, J., Wang, L.E., Ma, H., Liu, J., Zhao, Y., Tang, J., Chen, X., Dai, J., Wei, Q., Zhang, C. and Shen, H., 2012. Serum microRNA profiling and breast cancer risk: the use of miR-484/191 as endogenous controls. *Carcinogenesis*, **33**(4), pp. 828-34.

Huang, Z., Huang, D., Ni, S., Peng, Z., Sheng, W. and Du, X., 2010. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int J Cancer*, **127**(1), pp. 118-26.

Hunter, M.P., Ismail, N., Zhang, X., Aguda, B.D., Lee, E.J., et al., 2008. Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One*, **3**, pp. e3694.

Husemann, Y., Geigl, J.B., Schubert, F., Musiani, P., Meyer, M., Burghart, E., et al., 2008. Systemic spread is an early step in breast cancer. *Cancer Cell*, **13**, pp. 58–68.

Huw, L.Y., O'Brien, C., Pandita, A., Mohan, S., Spoerke, J.M., Lu, S., Wang, Y., Hampton, G.M., Wilson, T.R., Lackner, M.R., 2013. Acquired PIK3CA amplification causes resistance to selective phosphoinositide 3-kinase inhibitors in breast cancer. *Oncogenesis*, **23**, pp. 2:e83.

Iafrate, A.J., et al., 2004. Detection of large-scale variation in the human genome. *Nature Genetics*, **36**, pp. 949–951.

Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., et al., 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*, **65**, pp. 7065–7070.

Isakoff, S.J., Engelman, J.A., Irie, H.Y., et al., 2005. Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer Res*, **65**(23), pp. 10992–11000.

Isler, J.A., Vesterqvist, O.E. and Burczynski, M.E. 2007. Analytical validation of genotyping assays in the biomarker laboratory. *Pharmacogenomics*, **8**, pp. 353–368.

Jahr, S., Hentze, H., Englisch, S., et al., 2001. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res*, **61**, pp. 1659–65.

Jarry, J., Schadendorf, D., Greenwood, C., Spatz, A. and van Kempen, L.C., 2014. The validity of circulating microRNAs in oncology: five years of challenges and contradictions. *Mol Oncol*, **8**, pp. 819–29.

Jelovac, D., Beaver, J.A., Balukrishna, S., Wong, H.Y., Toro, P.V., Cimino-Mathews, A., Argani, P., Stearns, V., Jacobs, L., VanDenBerg, D., Kessler, J., Jeter, S., Park, B.H. and Wolff, A.C., 2014. A PIK3CA mutation detected in plasma from a patient with synchronous primary breast and lung cancers. *Human Pathology*, **45**, pp. 880–883

Jensen, J.D., Knoop, A., Laenkholm, A.V., Grauslund, M., Jensen, M.B., et al., 2012. PIK3CA mutations, PTEN, and pHER2 expression and impact on outcome in HER2-positive early-stage breast cancer patients treated with adjuvant chemotherapy and trastuzumab. *Ann Oncol* **23**, pp. 2034–2042.

Jiang, W.W., Zahurak, M., Goldenberg, D., et al., 2006. Increased plasma DNA integrity index in head and neck cancer patients. *Int J Cancer*, 119, pp. 2673–6.

Jiang, Y.Z., Yu, K.D., Bao, J., Peng, W.T. and Shao, Z.M., 2014. Favorable prognostic impact in loss of TP53 and PIK3CA mutations after neoadjuvant chemotherapy in breast cancer. *Cancer Res*, **74**, pp. 3399e407.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D. and Slack, F.J., 2005. RAS is regulated by the let-7 microRNA family. *Cell,* **120**, pp. 635–647.

Jonkers, J., Meuwissen, R., van der Gulden., H., Peterse, H., van der Valk, M. and Berns, A., 2001. Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet*, **29**, pp. 418–25.

Jung, E.J., Santarpia, L., Kim, J., Esteva, F.J., Moretti, E., Buzdar, A.U., Di Leo, A., Le, X.F., Bast, R.C., Park, S.T., Pusztai, L. and Calin, G.A., 2012. Plasma microRNA 210 levels correlate with sensitivity to trastuzumab and tumor presence in breast cancer patients. *Cancer*, **118**(10), pp. 2603-14.

Kalinsky, K., Jacks, L.M., Heguy, A., et al., 2009. PIK3CA mutation associates with improved outcome in breast cancer. *Clin Cancer Res*, **15**, pp. 5049–59

Kang, J.H., Kim, S.J., Noh, D.Y., Choe, K.J., Lee, E.S. and Kang, H.S., 2001. The timing and characterization of p53 mutations in in progression from atypical ductal hyperplasia to invasive lesions in the breast cancer. *J Mol Med (Berl)*, **79**(11), pp. 648-55.

Kang, S., Bader, A.G. and Vogt, P.K., 2005. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc Natl Acad Sci USA*, **102**(3), pp. 802–807.

Karakas, B., Bachman, K. and Park, B., 2006. Mutation of the PIK3CA oncogene in human cancers. *Br J Cancer*, **94**, pp. 455–9.

Karrison, T. G., Ferguson, D. J., and Meier, P., 1999. Dormancy of mammary carcinoma after mastectomy. *Journal of the National Cancer Institute*, **91**(1), pp. 80–85.

Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M. and Ruvkun, G., 2004. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc Natl Acad Sci U S A*, **101**(1), pp. 360-365.

Kim, H.S., Yom, C.K., Kim, H.J., Lee, J.W., Sohn, J.H, Kim, J.H., Park, Y.L. and Ahn, S.H., 2010. Overexpression of p53 is correlated with poor outcome in premenopausal women with breast cancer treated with tamoxifen after chemotherapy. *Breast Cancer Res. Treat.* **121**, pp. 777–788.

Kim, D.J., Linnstaedt, S., Palma, J., Park, J.C., Ntrivalas, E., et al., 2012. Plasma components affect accuracy of circulating cancer-related microRNA quantitation. *J Mol Diagn*, **14**, pp. 71–80.

Kinde, I, Wu., J, Papadopoulos, N., Kinzler, K.W. and Vogelstein, B., 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A*, **108**, pp. 9530e9535.

Kirschner, M.B., Edelman, J.J., Kao, S.C., Vallely, M.P., van Zandwijk, N. and Reid, G., 2013. The Impact of Hemolysis on Cell-Free microRNA Biomarkers. *Front Genet*, **4**, pp. 94.

Kloosterman, W.P. and Plasterk, R.H., 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell*, **11**, pp. 441-450.

Ko, L.J. and Prives, C., 1996. p53: puzzle and paradigm. *Genes Dev*, **10** (9), 1054–1072.

Kodahl, A.R., Lyng, M.B., Binder, H., Cold, S., Gravgaard, K., Knoop, A.S. and Ditzel, H.J., 2014. Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: a case control study. *Mol Oncol*, **8**(5), pp. 874-83.

Kohlmann, A., Klein, H.U., Weissmann, S., Bresolin, S., Chaplin, T., Cuppens, H., Haschke-Becher, E., Garicochea, B., Grossmann, V., Hanczaruk, B., Hebestreit, K., Gabriel, C., Iacobucci, I., Jansen, J.H., te Kronnie, G., van de Locht, L., Martinelli, G., McGowan, K., Schweiger, M.R., Timmermann, B., Vandenberghe, P., Young, B.D., Dugas, M. and Haferlach, T., 2011. The Interlaboratory RObustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. *Leukemia*, **25**(12), pp. 1840-8.

Kosaka, N., Iguchi, H. and Ochiya, T., 2010. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Sci*, **101**, pp. 2087–2092.

Kwong, A., Ng, E.K., Tang, E.Y., Wong, C.L., Law, F.B., Leung C.P., et al., 2011. A novel de novo BRCA1 mutation in a Chinese woman with early onset breast cancer. *Familial cancer*, **10**(2), pp. 233-7.

Lam, N.Y., Rainer, T.H., Chiu, R.W., Joynt, G.M. and Lo, Y.M., 2004. Plasma mitochondrial DNA concentrations after trauma. *Clin Chem*, **50**, pp. 213-216.

Lawrie, C.H., Gal, S., Dunlop, H.M., et al., 2008. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol*, **141**, pp. 672–5.

Lee, R.C., Feinbaum, R.L. and Ambros, V., 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, pp. 843–854.

Lee, C.S., Taib, N.A., Ashrafzadeh, A., Fadzli, F., Harun, F., Rahmat, K., Hoong, S.M., Abdul-Rahman, P.S. and Hashim, O.H., 2016. Unmasking Heavily O-Glycosylated Serum Proteins Using Perchloric Acid: Identification of Serum Proteoglycan 4 and

Protease C1 Inhibitor as Molecular Indicators for Screening of Breast Cancer. *PLoS One*, **11**(2), pp. e0149551

Lehmann, B.D. and Pietenpol, J.A., 2012. Targeting mutant p53 in human tumors. *J Clin Oncol*, **30**, pp. 3648–50.

Leon, S.A., Shapiro, B., Sklaroff, D.M. and Yaros, M.J., 1977. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res*, **37**, pp. 646–50.

Levine, A.J., Momand, J. and Finlay, C.A., 1991. The p53 tumour suppressor gene. *Nature*, **351** (6326), pp. 453–456.

Levine, A.J., 1997. p53, the cellular gatekeeper for growth and division. *Cell*, **88**, pp. 323–31.

Levine, D.A., Bogomolniy, F., Yee, C.J., Lash, A., Barakat, R.R., Borgen, P.I., et al., 2005. Frequent mutation of the PIK3CA gene in ovarian and breast cancer. *Clin Cancer Res*, **11**, pp. 2875–8.

Li, H., Zhu, R., Wang, L., Zhu, T., Li, Q., Chen, Q., Wang, H., Zhu, H., 2010. PIK3CA mutations mostly begin to develop in ductal carcinoma of the breast. *Exp Mol Pathol*, **88**(1), pp. 150–155.

Li, D., Zhao, Y., Liu, C., Chen, X., Qi, Y., Jiang, Y., Zou, C., Zhang, X., Liu, S., Wang, X., et al., 2011. Analysis of MiR-195 and MiR-497 expression, regulation and role in breast cancer. *Clin Cancer Res*, **17**(7), pp. 1722-1730.

Li, Q., Eades, G., Yao, Y., Zhang, Y. and Zhou, Q., 2014. Characterization of a stem-like subpopulation in basal-like ductal carcinoma in situ (DCIS) lesions. *J Biol Chem*, **289**, pp. 1303–12.

Liang, Y., Ridzon, D., Wong, L. and Chen, C., 2007. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, **8**, pp. 166.

Ligresti, G., Militello, L., Steelman, L.S., Cavallaro, A., Basile, F., Nicoletti, F., Stivala, F., McCubrey, J.A. and Libra, M., 2009. PIK3CA mutations in human solid tumors: role in sensitivity to various therapeutic approaches. *Cell Cycle*, **8**(9):1352-8.

Linzer, D.I. and Levine, A.J., 1979. Characterization of a 54K Dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell*, 17, pp. 43–52.

Liu, P., Cheng, H., Roberts, T.M. and Zhao, J.J., 2009. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov*, **8**, pp. 627–44.

Liu, W., Sun, J., Li, G., Zhu, Y., Zhang, S., Kim, S.T., Sun, J., Wiklund, F., Wiley, K., Isaacs, S.D., Stattin, P., Xu, J., Duggan, D., Carpten, J.D., Isaacs, W.B., Gronberg, H., Zheng, S.L. and Chang, B.L., 2009. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res,* **69**, pp. 2176-2179.

Liu, R., Liao, J., Yang, M., Shi, Y., Peng, Y., Wang, Y., Pan, E., Guo, W., Pu, Y. and Yin, L., 2012. Circulating miR-155 expression in plasma: a potential biomarker for early diagnosis of esophageal cancer in humans. *J Toxicol Environ Health A,* **75**(18):1154-62.

Lo, Y.M., Zhang, J., Leung, T.N., Lau, T.K., Chang, A.M. and Hjelm, N.M., 1999. Rapid clearance of fetal DNA from maternal plasma. *Am J Hum Genet,* **64**(1), pp. 218–224.

Lodes, M.J., Caraballo, M., Suciu, D., Munro, S., Kumar, A., et al., 2009. Detection of cancer with serum miRNAs on an oligonucleotide microarray. *PLoS One*, **4**, pp. e6229.

Loeb, L.A. and Christians, F.C., 1996. Multiple mutations in human cancers. *Mutat. Res.*, **350**, pp. 279*–286.*

Loi, S., Haibe-Kains, B., Majjaj, S., et al., 2010. PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor-positive breast cancer. *Proc Natl Acad Sci U S A,* **107**(22), pp. 10208–10213.

Loi, S., Michiels, S., Lambrechts, D., Fumagalli, D., Claes, B., et al., 2013. Somatic mutation profiling and associations with prognosis and trastuzumab benefit in early breast cancer. *J Natl Cancer Inst,* **105**, pp. 960– 967.

Lucito, R., Suresh, S., Walter, K., Pandey, A., Lakshmi, B., Krasnitz, A., Sebat, J., Wigler, M., Klein, A.P., Brune, K., Palmisano, E., Maitra, A., Goggins, M. and Hruban, R.H., 2007. Copy-number variants in patients with a strong family history of pancreatic cancer. *Cancer Biol Ther*, **6**, pp. 1592-1599.

Luo, J., Zhao, Q., Zhang, W., Zhang, Z., Gao, J., Zhang, C., et al., 2014. A novel panel of microRNAs provides a sensitive and specific tool for the diagnosis of breast cancer. *Mol Med Rep,* **10**(2), pp. 785–91.

Ma, L., Teruya-Feldstein, J., Weinberg, R.A., 2007. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature,* **449**, pp. 682-688.

Ma, N., Zhang, W., Qiao, C., Luo, H., Zhang, X., Liu, D., Zang, S., Zhang, L. and Bai, J., 2016. The Tumor Suppressive Role of MiRNA-509-5p by Targeting FOXM1 in Non-Small Cell Lung Cancer. *Cell Physiol Biochem*, **38**(4), pp. 1435-46.

Madjar H., 2010. Role of Breast Ultrasound for the Detection and Differentiation of Breast Lesions. *Breast Care (Basel).*, **5(**2), pp. 109-114.

Mangolini, A., Ferracin, M., Zanzi, M.V., Saccenti, E., Ebnaof, S.O., Poma, V.V., Sanz, J.M., Passaro, A., Pedriali, M., Frassoldati, A., Querzoli, P., Sabbioni, S., Carcoforo, P., Hollingsworth, A. and Negrini, M., 2015. Diagnostic and prognostic microRNAs in the serum of breast cancer patients measured by droplet digital PCR. *Biomarker Research*, **3**, pp. 12.

Mao, C., Yang, Z.Y., Hu, X.F., Chen, Q. and Tang, J.L., 2012. PIK3CA exon 20 mutations as a potential biomarker for resistance to anti-EGFR monoclonal antibodies in KRAS wild-type metastatic colorectal cancer: a systematic review and meta-analysis. *Ann Oncol*, **23**(6), pp. 1518–25.

Mar-Aguilar, F., Mendoza-Ramirez, J.A., Malagon-Santiago, I., Espino-Silva, P.K., Santuario-Facio, S.K., Ruiz-Flores, P., et al., 2013. Serum circulating microRNA profiling for identification of potential breast cancer biomarkers. *Dis Markers*, **34**, pp. 163–9.

Maric, P., Ozretic, P., Levanat, S.S., Oreskovic, S., Antunac, K. and Beketic-Oreskovic, L., 2011. Tumor markers in breast cancer—evaluation of their clinical usefulness. *Coll Antropol*, **35**(1), pp. 241–7.

Maruyama, N., Miyoshi, Y., Taguchi, T., Tamaki, Y., Monden, M. and Noguchi S., 2007. Clinicopathologic analysis of breast cancer with PIK3CA mutations in Japanese women. *Clin Cancer Res*, **13**, pp. 408–14.

Matamala, N., Vargas, M.T., González-Campora, R., Minambres, R., Arias, J.I., Menendez, P., Andres-Leon, E., Gomez-Lopez, G., Yanowsky, K., Calvete-Candenas, J., Inglada-Perez, L., Martinez-Delgado, B. and Benitez, J., 2015. Tumor microRNA expression profiling identifies circulating microRNAs for early breast cancer detection. *Clin Chem.*, **61**(8), pp. 1098-106.

Maxmen, A., 2012. The hard facts. *Nature*, **485**, pp. S50–1.

McDonald, J.S., Milosevic, D., Reddi, H.V., Grebe, S.K. and Algeciras-Schimnich, A., 2011. Analysis of circulating microRNA: preanalytical and analytical challenges. *Clin Chem*, **57**, pp. 833–840.

Meltzer, A., 1990. Dormancy and breast cancer. *Journal of Surgical Oncology*, **43**(3), pp. 181–188.

Mestdagh, P., Hartmann, N., Baeriswyl, L., Andreasen, D., Bernard, N., Chen, C., et al., 2014. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat Methods*, **11**, pp. 809–15.

Milbury, C.A., Li, J. and Makrigiorgos, G.M., 2009. PCR-based methods for the enrichment of minority alleles and mutations. *Clin Chem*, **55**, pp. 4632e4640

Milbury, C.A., Correll, M., Quackenbush, J., Rubio, R. and Makrigiorgos, G.M., 2012. COLD-PCR enrichment of rare cancer mutations prior to targeted amplicon resequencing. *Clin Chem*, **58**, pp. 580e589

Miron, A., Varadi, M., Carrasco, D., Li, H., Luongo, L., Kim, H.J., et al., 2010. PIK3CA mutations in in situ and invasive breast carcinomas. *Cancer Res*, **70**, pp. 5674–8.

Mishra, S., Srivastava, A.K., Suman, S., Kumar, V. and Shukla, Y., 2015. Circulating miRNAs revealed as surrogate molecular signatures for the early detection of breast cancer. *Cancer Lett*, **369**(1), pp. 67-75.

Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., et al., 2008. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*, **105**, pp. 10513–10518.

Moon, P.G., Lee, J.E., Cho, Y.E., Lee, S.J., Jung, J.H., Chae, Y.S., Bae, H.I., Kim, Y.B., Kim, I.S., Park, H.Y. and Baek, M.C., 2015. Identification of Developmental Endothelial Locus-1 on Circulating Extracellular Vesicles as a Novel Biomarker for Early Breast Cancer Detection. *Clin Cancer Res*, **22**(7), pp.1757-66.

Morganella, S., Alexandrov, L.B., Glodzik, D., Zou, X., Davies, H., Staaf, J., Sieuwerts, A.M., Brinkman, A.B., Martin, S., Ramakrishna, M., Butler, A., Kim, H.Y., Borg, A., Sotiriou, C., Futreal, P.A., Campbell, P.J., Span, P.N., Van Laere, S., Lakhani, S.R., Eyfjord, J.E., Thompson, A.M., Stunnenberg, H.G., van de Vijver, M.J., Martens, J.W., Borresen-Dale, A.L., Richardson, A.L., Kong, G., Thomas, G., Sale, J., Rada, C., Stratton, M.R., Birney, E. and Nik-Zainal, S., 2016. The topography of mutational processes in breast cancer genomes. *Nat Commun*, **7**, pp. 11383.

Muller, C.I., Miller, C.W., Hofmann, W.K., et al., 2007. Rare mutations of the PIK3CA gene in malignancies of the hematopoietic system as well as endometrium, ovary, prostate and osteosarcomas, and discovery of a PIK3CA pseudogene. *Leuk Res*, **31**, pp. 27–32.

Murray, C., 1995. Tumour dormancy: not so sleepy after all. *Nature Medicine*, **1**(2), pp. 117–118.

My cancer genome (www.mycancergenome.org).

Nadji, M., Gomez-Fernandez, C., Ganjei-Azar, P. & Morales, A.R., 2005. Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5,993 breast cancers**.** *Am J Clin Pathol.* **123**(1), pp. 21-7.

Nakauchi, C., Kagara, N., Shimazu, K., Shimomura, A., Naoi, Y., Shimoda, M., Kim, S.J. and Noguchi, S., 2016. Detection of TP53/PIK3CA Mutations in Cell-Free Plasma DNA From Metastatic Breast Cancer Patients Using Next Generation Sequencing. *Clin Breast Cancer*, pii: S1526-8209(16), pp. 30100-8.

Narayan, A., Carriero, N.J., Gettinger, S.N., Kluytenaar, J., Kozak, K.R., Yock, T.I., Muscato, N.E., Ugarelli, P., Decker, R.H. and Patel, A.A, 2012. Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Res*, **72**, pp. 3492-3498

Nattinger, A.B., 2010. In the clinic. Breast cancer screening and prevention. *Ann. Intern. Med.*, **152**(7), ITC41.

Nelson, H.D., Tyne, K., Naik, A., et al., 2009. Screening for breast cancer: U.S. preventive services task force recommendation statement. *Ann Intern Med*, **151**, pp. 716-726.

Ng, P.C., Levy, S., Huang, J., Stockwell, T.B., Walenz, B.P., Li, K., Axelrod, N., Busam, D.A., Strausberg, R.L. and Venter, J.C., 2008. Genetic variation in an individual human exome. *PLoS Genet*, **4**, pp. e1000160.

Ng, E.K., Chong, W.W., Jin, H., Lam, E.K., Shin, V.Y., et al., 2009. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut*, **58**, pp. 1375–1381.

Ng, E.K., Li, R., Shin, V.Y., Jin, H.C., Leung, C.P., Ma, E.S., et al., 2013. Circulating micro‐RNAs as specific biomarkers for breast cancer detection. *PLoS One*, **8**, pp. e53141.

Nik-Zainal, S. et al., 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell*, **149**, pp. 979–993.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., Van Loo, P., Ju, Y.S. Smid, M., Brinkman, A.B., Morganella, S., Aure, M.R., Lingjaerde, O.C., Langerod, A., Ringner, M., Ahn, S.M., Boyault, S., Brock, J.E., Broeks, A., Butler, A., Desmedt, C., Dirix, L., Dronov, S., Fatima, A., Foekens, J.A., Gerstung, M., Hooijer, G.K., Jang, S.J., Jones, D.R., Kim, H.Y., King, T.A., Krishnamurthy, S., Lee, H.J., Lee, J.Y., Li, Y., McLaren, S., Menzies, A., Mustonen, V., O'Meara, S., Pauporte, I., Pivot, X., Purdie, C.A., Raine, K., Ramakrishnan, K., Rodriguez-Gonzalez, F.G., Romieu, G., Sieuwerts, A.M. Simpson P.T., Shepherd R., Stebbings L., Stefansson O.A., Teague J., Tommasi S., Treilleux I., Van den Eynden, G.G., Vermeulen, P., Vincent-Salomon, A., Yates, L., Caldas, C., van't Veer, L., Tutt, A., Knappskog, S., Tan, B.K., Jonkers, J., Borg, A., Ueno, N.T., Sotiriou, C., Viari, A., Futreal, P.A., Campbell, P.J., Span, P.N., Van Laere, S., Lakhani, S.R., Eyfjord, J.E., Thompson, A.M., Birney, E., Stunnenberg, H.G., van de Vijver, M.J., Martens, J.W., Borresen-Dale, A.L., Richardson, A.L., Kong, G., Thomas, G. and Stratton, M.R., 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, **534**(7605), pp. 47-54.

O'Day, E. and Lal, A., 2010. MicroRNAs and their target gene networks in breast cancer. Breast *Cancer Res*, **12**: pp. 201.

Okkenhaug, K., 2013. Signaling by the phosphoinositide 3-kinase family in immune cells. *Annu Rev Immunol*, **31**, pp. 675–704.

OligoEvaluator™ (www.sigmaaldrich.com).

Olivier, M. Langerod, A., Carrieri, P., Bergh, J., Klaar, S., Eyfjord, J., Theillet, C., Rodriguez, C., Lidereau, R., Bieche, I., Varley, J., Bignon, Y., Uhrhammer, N., Winqvist, R., Jukkola-Vuorinen, A., Niederacher, D., Kato, S., Ishioka, C., Hainaut, P. and Borresen-Dale, AL., 2006. The clinical value of somatic TP53 gene mutations in 1794 patients with breast cancer. Clin. Cancer Res, **12**(4), pp. 1157-67.

Oren, M., 1999. Regulation of the p53 tumor suppressor protein. *J. Biol. Chem*, **274** (51), pp. 36031–36034.

Oshiro, C., Kagara, N., Naoi, Y., Shimoda, M., Shimomura, A., Maruyama, N., et al., 2015. PIK3CA mutations in serum DNA are predictive of recurrence in primary breast cancer patients. *Breast Cancer Res Treat*, **150**(2), pp. 299–307.

Page, K., Powles, T., Slade, M.J., Tamburo de Bella, M., Walker, R.A., Coombes, R.C. and Shaw, J.A., 2006. The importance of careful blood processing in isolation of cell-free DNA. *Ann N Y Acad Sci*, **1075**, pp. 313–317.

Page, K., Hava, N., Ward, B., Brown, J., Guttery, D.S., Ruangpratheep, C., Blighe, K., Sharma, A., Walker, R.A., Coombes, R.C. and Shaw JA., 2011. Detection of HER2

amplification in circulating free DNA in patients with breast cancer. *Br J Cancer.*, **104**(8):1342-8.

Page, K., Guttery, D.S., Zahra, N., Primrose, L., Elshaw, S.R., Pringle, J.H., et al., 2013. Influence of plasma processing on recovery and analysis of circulating nucleic acids. *PloS one*, **8**, pp. e77963.

Palmieri, C., Cleator, S., Kilburn, L.S., Kim, S.B., Ahn, S.H., Beresford, M., Gong, G., Mansi, J., Mallon, E., Reed, S., Mousa, K., Fallowfield, L., Cheang, M., Morden, J., Page, K., Guttery, D.S., Rghebi, B., Primrose, L., Shaw, J.A., Thompson, A.M., Bliss, J.M. and Coombes, RC., 2014. NEOCENT: a randomised feasibility and translational study comparing neoadjuvant endocrine therapy with chemotherapy in ER-rich postmenopausal primary breast cancer. *Breast Cancer Res Treat*, **148**(3), pp. 581-90.
Pantel, K. and Speicher, M.R., 2016. The biology of circulating tumor cells. *Oncogene,* **35**, pp. 1216-1224.

Parant, J.M. and Lozano, G., 2003. Disrupting TP53 in mouse models of human cancers. *Hum. Mutat.*, **21**, pp. 321–326.

Park, N.J., Zhou, H., Elashoff, D., Henson, B.S., Kastratovic, D.A., et al., 2009. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. *Clin Cancer Res*, **15**, pp. 5473–5477.

Parkinson, D.R., Dracopoli, N., Petty, B.G., Compton, C., Cristofanilli, M., Deisseroth, A., et al., 2012. Considerations in the development of circulating tumor cell technology for clinical use. *J Translat Med*, **10**, pp. 138.

Perou, C., et al, 2000. Molecular portraits of human breast tumours. *Nature*, **406**(6797), pp. 747–752

Petricoin, E.F., Ornstein, D.K., Paweletz, C.P., Ardekani, A., Hackett, P.S., Hitt, B.A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C.B., Levine, P.J., Linehan, W.M., Emmert-Buck, M.R., Steinberg, S.M., Kohn, E.C. and Liotta, L.A., 2002. Serum proteomic patterns for detection of prostate cancer. *Journal of the National Cancer Institute,* **94**(20), pp. 1576–1578.

Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P., et al., 2007. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat*, **28**, pp. 622–9.

Petrij-Bosch, A., Peelen, T., van Vliet, M., van Eijk, R., Olmer, R., Drusedau, M., Hogervorst, F.B., Hageman, S., Arts, P.J., Ligtenberg, M.J., Meijers-Heijboer, H.,

Klijn, J.G., Vasen, H.F., Cornelisse, C.J., van 't Veer, L.J., Bakker, E., van Ommen, G.J. and Devilee, P., 1997. *BRCA1* genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat Genet*, **17**, pp. 341-345.

Pfundheller, H.M., Sorensen, A.M., Lomholt, C., Johansen, A.M., Koch, T. and Wengel, J., 2005. Locked nucleic acid synthesis. *Methods Mol Biol*, **288**, 127e146.

Philp, A.J., Campbell, I.G., Leet, C., Vincan, E., Rockman, S.P., Whitehead, R.H., et al., 2001. The phosphatidylinositol 3'-kinase p85alpha gene is an oncogene in human ovarian and colon tumours. *Cancer Res*, **61**, pp. 7426–9.

Pigati, L., Yaddanapudi, S.C., Iyengar, R., Kim, D.J., Hearn, S.A., Danforth, D., Hastings, M.L. and Duelli, D.M., 2010. Selective release of microRNA species from normal and malignant mammary epithelial cells. *PLoS One*, **5**, pp. e13515.

Platt, F.M., Hurst, C.D., Taylor, C.F., et al., 2009. Spectrum of phosphatidylinositol 3-kinase pathway gene alterations in bladder cancer. *Clin Cancer Res*, **15**, pp. 6008–6017.

Pritchard, C.C., Kroh, E., Wood, B., Arroyo, J.D., Dougherty, K.J., et al., 2012. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)*, **5**, pp. 492–497.

PubMed NCBI website (www.ncbi.nlm.nih.gov/pubmed).

Ramirez, J.L., Taron, M., Balana, C., Sarries, C., Mendez, P., de Aguirre, I., Nunez, L., Roig, B., Queralt, C., Botia, M. and Rosell, R., 2003 a. Serum DNA as a tool for cancer patient management. *Rocz Akad Med Bialymst*, **48**, pp. 34-41.

Ramirez, J.L., Sarries, C., de Castro, P.L., et al., 2003 b. Methylation patterns and K-ras mutations in tumor and paired serum of resected non-small-cell lung cancer patients. *Cancer Lett*, **193**, pp. 207–16.

Redon, R., et al., 2006. Global variation in copy number in the human genome. *Nature* , **444**, pp. 444–454.

Ren, Z.J., Nong, X.Y., Lv, Y.R., Sun, H.H., An, P.P., Wang, F., Li, X., Liu, M. and Tang, H., 2014. Mir-509-5p joins the Mdm2/p53 feedback loop and regulates cancer cell growth. *Cell Death Dis*, **5**, pp. e1387.

Rosell, R., Moran, T., Queralt, C., Porta, R., Cardenal, F., Camps, C., et al., 2009. Screening for epidermal growth factor receptor mutations in lung cancer. *The New England journal of medicine*, **361**, pp. 958-67.

Roth, C., Rack, B., Muller, V., Janni, W., Pantel, K. and Schwarzenbach, H., 2010. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. *Breast Cancer Res*, **12**(6), pp. R90.

Saal, L.H., Holm, K., Maurer, M., Memeo, L., Su, T., Wang, X., et al., 2005. PIK3CAmutations correlate with hormone receptors, nodemetastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Res*, **65**, pp. 2554–9.

Sabine, V.S., Crozier, C., Drake, C., Piper, T., van de Velde, C.J., Hasenburg, A., et al., 2012. PIK3CA mutations are linked to PgR expression: a tamoxifen exemestane adjuvant multinational (TEAM) pathology study. *Cancer Res*, **72**, pp. 1–5.

Sabine, V.S., Crozier, C., Brookes, C.L., Drake, C., Piper, T., van de Velde, C.J., et al., 2014. Mutational analysis of PI3K/AKT signaling pathway in tamoxifen exemestane adjuvant multinational pathology study. *J Clin Oncol*, **32**(27), pp. 2951–8.

Salem, D.S., Kamal, R.M., Mansour, S.M., Salah, L.A. & Wessam, R., 2013. Breast imaging in the young: the role of magnetic resonance imaging in breast cancer screening, diagnosis and follow-up. *J Thorac Dis*., **5** (Suppl 1), pp. S9-S18.

Samuels, Y., Wang, Z., Bardelli, A., Siliman, N., Ptak, J., Szabo, S., et al., 2004. High frequency of mutations of the PIK3CA gene in human cancers. *Science*, **304**, pp. 554.

Samuels, Y., Diaz, L.A. Jr, Schmidt-Kittler, O., Cummins, J.M., Delong, L., Cheong, I., Rago, C., Huso, D.L., Lengauer, C., Kinzler, K.W., Vogelstein, B. and Velculescu, V.E., 2005. Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell*, **7**(6), pp. 561–573

Samuels, Y. and Ericson, K., 2006. Oncogenic PI3K and its role in cancer. *Curr Opin Oncol*, **18**, pp. 77–82.

Schetter, A.J., Leung, S.Y., Sohn, J.J., Zanetti, K.A., Bowman, E.D., et al., 2008. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA*, **299**, pp. 425–436.

Schlechte, H.H., Stelzer, C., Weickmann, S., Fleischhacker, M. and Schulze, G., 2004. TP53 gene in blood plasma DNA of tumor patients. *Ann N Y Acad Sci*, **1022**, pp. 61-9.

Schonleben, F., Qiu, W., Remotti, H.E., et al., 2008. PIK3CA, KRAS, and BRAF mutations in intraductal papillary mucinous neoplasm/carcinoma (IPMN/C) of the pancreas. *Langenbecks Arch Surg*, **393**, pp. 289–296.

Schrauder, M.G., Strick, R., Schulz-Wendtland, R., Strissel, P.L., Kahmann, L., Loehberg, C.R., Lux, M.P., Jud, S.M., Hartmann, A., Hein, A., Bayer, C.M., Bani, M.R., Richter, S., Adamietz, B.R., Wenkel, E., Rauh, C., Beckmann, M.W. and Fasching, P.A., 2012. Circulating micro-RNAs as potential blood-based markers for early stage breast cancer detection. *PLoS One*, **7**(1), pp. e29770.

Schwarzenbach, H., Milde-Langosch, K., Steinbach, B., Muller, V., Pantel, K., 2012a. Diagnostic potential of PTEN-targeting miR-214 in the blood of breast cancer patients. *Breast Cancer Res Treat*, **134**(3):933-41.

Schwarzenbach, H., Eichelser, C., Kropidlowski, J., Janni, W., Rack, B. and Pantel, K., 2012b. Loss of heterozygosity at tumor suppressor genes detectable on fractionated circulating cell-free tumor DNA as indicator of breast cancer progression. *Clin Cancer Res*, **18**(20), pp. 5719-30.

Schwarzenbach, H., 2013. Circulating nucleic acids as biomarkers in breast cancer. *Breast Cancer Res*, **15**(5), pp. 211.

Sebat, J., et al., 2004. Large-scale copy number polymorphism in the human genome. *Science*, **305**, pp. 525–528.

Shaw, J.A., Brown, J., Coombes, R.C., Jacob, J., Payne, R., Lee, B., Page, K., Hava, N. and Stebbing, J., 2011. Circulating tumor cells and plasma DNA analysis in patients with indeterminate early or metastatic breast cancer. *Biomarkers Med*, **5**, pp. 87–91.

Shaw, J.A., Page, K., Blighe, K., Hava, N., Guttery, D., Ward, B., Brown, J., Ruangpratheep, C., Stebbing, J., Payne, R., Palmieri, C., Cleator, S., Walker, R.A. and Coombes, R.C., 2012. Genomic analysis of circulating cell-free DNA infers breast cancer dormancy. *Genome Res*, **22**(2), pp. 220-31.

Sigal, A. and Rotter, V., 2000. Oncogenic mutations of the p53 tumor suppressor. The demons of the guardian of the genome. *CancerRes*, **60**, pp. 6788–93.

Silva, J.M., Dominguez, G., Garcia, J.M., Gonzalez, R., Villanueva, M.J., Navarro, F., Provencio, M., San Martin, S., Espana, P. and Bonilla, F., 1999. Presence of tumor DNA in plasma of breast cancer patients: clinicopathological correlations. *Cancer Res*, **59**, pp. 3251-3256.

Singh, R. and Mo, Y.Y., 2013. Role of microRNAs in breast cancer. *Cancer Biol Ther*, **14**, pp. 201–212.

Song, B., Wang, C., Liu, J., Wang, X., Lv, L., Wei, L., Xie, L., Zheng, Y. and Song, X., 2010. MicroRNA-21 regulates breast cancer invasion partly by targeting tissue inhibitor of metalloproteinase 3 expression. *J Exp Clin Cancer Res*, **29**, pp. 29.

Song, M.Y., Pan, K.F., Su, H.J., Zhang, L., Ma, J.L., Li, J.Y., Yuasa, Y., Kang, D., Kim, Y.S. and You, W.C., 2012. Identification of serum microRNAs as novel non-invasive biomarkers for early detection of gastric cancer. *PLoS One*, **7**(3), pp. e33608.

Souglakos, J., Philips, J., Wang, R., Marwah, S., Silver, M., Tzardi, M., et al., 2009. Prognostic and predictive value of commonmutations for treatment response and survival in patients with metastatic colorectal cancer. *Br J Cancer*, **101**, pp. 465–72.

Sourvinou, I.S., Markou, A. and Lianidou, ES., 2013. Quantification of circulating miRNAs in plasma: effect of preanalytical and analytical parameters on their isolation and stability. *J Mol Diagn*, **15**, pp. 827–34.

Stankiewicz, P. and Lupski, J.R., 2010. Structural variation in the human genome and its role in disease. *Annu Rev Med*, **61**, pp. 437-455.

Stemke-Hale, K., Gonzalez-Angulo, A.M., Lluch, A., Neve, R.M., Kuo, W.L., et al., 2008. An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res*, **68**, pp. 6084–6091.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., Yates, L.R., Papaemmanuil, E., Beare, D., Butler, A., Cheverton, A., Gamble, J., Hinton, J., Jia, M., Jayakumar, A., Jones, D., Latimer, C., Lau, K.W., McLaren, S., McBride, D.J., Menzies, A., Mudie, L., Raine, K., Rad, R., Chapman, M.S., Teague, J., Easton, D., Langerod, A., Lee, M.T., Shen, C.Y., Tee, B.T., Huimin, B.W., Broeks, A., Vargas, A.C., Turashvili, G., Martens, J., Fatima, A., Miron, P., Chin, S.F., Thomas, G., Boyault, S., Mariani, O., Lakhani, S.R., van de Vijver, M., van't Veer, L., Foekens, J., Desmedt, C., Sotiriou, C., Tutt, A., Caldas, C., Reis-Filho, J.S., Aparicio, S.A., Salomon, A.V., Borresen-Dale, A.L., Richardson, A.L., Campbell, P.J., Futreal, P.A. and Stratton, M.R., 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature*, **486**(7403), pp. 400–404.

Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S.W., Tavare, S., Deloukas, P., Hurles, M.E. and Dermitzakis, E.T., 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, pp. 848-853.

Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. and Anker, P., 2001. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin Chim Acta*, **313**(1–2), pp. 139–142

Stuckrath, I., Rack, B., Janni, W., Jager, B., Pantel, K. and Schwarzenbach, H., 2015. Aberrant plasma levels of circulating mir-16, mir-107, mir-130a and mir-146a are associated with lymph node metastasis and receptor status of breast cancer patients. *Oncotarget*, **6**, pp. 13387.

Sudhakar, N., George Priya Doss, C., Thirumal Kumar, D., et al., 2015. Deciphering the impact of somatic mutations in exon 20 and exon 9 of PIK3CA gene in breast tumors among Indian women through molecular dynamics approach. *J Biomol Struct Dyn*, **34**, pp. 29–41.

Sun, X., Huang, J., Homma, T., et al., 2009. Genetic alterations in the PI3K pathway in prostate cancer. *Anticancer Res*, **29**, pp.1739–1743.

Sun, X., Qin, S., Fan, C., et al., 2013. Let-7: a regulator of the ERalpha signaling pathway in human breast tumors and breast cancer stem cells. *Oncol Rep*, **29**, pp. 2079–2087.

Taback, B. and Hoon, D.S., 2004 a. Circulating nucleic acids in plasma and serum: past, present and future. *Curr Opin Mol Ther*, **6**, pp. 273–8.

Taback, B. and Hoon, D.S., 2004 b. Circulating nucleic acids and proteomics of plasma/serum: clinical utility. *Ann N Y Acad Sci*, **1022**, pp. 1–8.

Tabar, L., Vitak, B., Chen, T.H., Yen, A.M., Cohen, A., Tot, T., Chiu, S.Y., Chen, S.L., Fann, J.C., Rosell, J., Fohlin, H., Smith, R.A. and Duffy, S.W., 2011. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*, **260**(3), pp. 658-663.

Taplin, S., Abraham, L., Barlow, W.E., Fenton, J.J., Berns, E.A.,Carney, P.A., et al., 2008. Mammography facility characteristics associated with interpretive accuracy of screening mammography. *Journal of the National Cancer Institute*, **100**(12), pp. 876–887.

Thean, L.F., Loi, C., Ho, K.S., Koh, P.K., Eu, K.W. and Cheah, P.Y., 2010. Genome-wide scan identifies a copy number variable region at 3q26 that regulates *PPM1L* in APC mutation-negative familial colorectal cancer patients. *Genes Chromosomes Cancer*, **49**, pp. 99-106.

Turchinovich, A., Weiz, L., Langheinz, A., et al., 2011. Characterization of extracellular circulating microRNA. *Nucleic Acids Res*, **39**, pp. 7223–33.

UCSC Genome Browser websites (genome.ucsc.edu).

Umetani, N., Giuliano, A.E., Hiramatsu, S.H., Amersi, F., Nakagawa, T., Martino, S. and Hoon, D.S., 2006a. Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol*, **24**, pp. 4270-4276.

Umetani, N., Kim, J., Hiramatsu, S., et al., 2006b. Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for repeats. *Clin Chem*, **52**, pp. 1062–9.

Van Der Vaart, M. and Pretorius PJ., 2008. Circulating DNA. Its origin and fluctuation. *Ann NY Acad Sci*, **1137**, pp. 18–26.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*., 3(7), RESEARCH0034.

Van Schooneveld, E., Wouters, M.C., Van der Auwera, I., Peeters, D.J., Wildiers, H., Van Dam, P.A., et al., 2012. Expression profiling of cancerous and normal breast tissues identifies microRNAs that are differentially expressed in serum from patients with (metastatic) breast cancer and healthy volunteers. *Breast Cancer Research*, **14**(1), pp. R34.

Varna, M., Bousquet, G., Plassa, L.F., Bertheau, P. and Janin, A., 2011. TP53 status and response to treatment in breast cancers. *Journal of Biomedicine and Biotechnology*, **2011**, pp. 284– 584.

Vogelstein, B., Lane, D. and Levine, A.J., 2000. Surfing the p53 network. *Nature*, **408**, pp. 307–10.

Voorhoeve, P.M., le Sage, C., Schrier, M., Gillis, A.J., Stoop, H., Nagel, R., Liu, Y.P., van Duijse, J., Drost, J., Griekspoor, A., et al., 2006. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell*, **124**, pp. 1169–1181.

Vousden, K.H., 2000. p53: death star. *Cell*, **103** (5), pp. 691–694.

Vousden, K.H., Lane, D.P., 2007. p53 in health and disease. *Nat. Rev. Mol. Cell Biol*, **8** (4), pp. 275–283.

Wang, B.G., Huang, H.Y., Chen, Y., et al., 2003. Increased plasma DNA integrity in cancer patients. *Cancer Res*, **63**, pp. 3966–8.

Wang, F., Zheng, Z., Guo, J. and Ding, X., 2010. Correlation and quantitation of microRNA aberrant expression in tissues and sera from patients with breast tumor. *Gynecol Oncol*, **119**(3), pp. 586-593.

Wang, H., Tan, G., Dong, L., Cheng, L., Li, K., Wang, Z. and Luo, H., 2012. Circulating MiR-125b as a marker predicting chemoresistance in breast cancer. *PLoS One*, **7**(4), pp. e34210.

Warnecke-Eberz, U., Chon, S.H., Holscher, A.H., Drebber, U. and Bollschweiler, E., 2015. Exosomal onco-miRs from serum of patients with adenocarcinoma of the esophagus: comparison of miRNA profiles of exosomes and matching tumor. *Tumour Biol*, **36**(6), pp. 4643-53.

Weber, G.F., 2008. Molecular mechanisms of metastasis. *Cancer Lett.*, **270**(2):181-90.

Weigel, M.T., and Dowsett, M., 2010. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocrine- Related Cancer*, **17**(4), pp. R245–R262.

Wolfson, B., Eades, G. and Zhou, Q., 2015. Adipocyte activation of cancer stem cell signaling in breast cancer. *World J Biol Chem*, **6**(2), pp. 39–47.

Wu, G., Xing, M., Mambo, E., Huang, X., Liu, J., Guo, Z., Chatterjee, A., Goldenberg, D., Gollin, S.M., Sukumar, S., Trink, B. and Sidransky, D., 2005. Somatic mutation and gain of copy number of *PIK3CA* in human breast cancer. *Breast Cancer Research*, **7**, pp. R609-R616.

Xing, F., Sharma, S., Liu, Y., Mo, Y.Y., Wu, K., Zhang, Y.Y., Pochampally, R., Martinez, L.A., Lo, H.W. and Watabe, K., 2015. miR- 509 suppresses brain metastasis of breast cancer cells by modulating RhoC and TNF-alpha. *Oncogene*, **34**, pp. 4890-4900.

Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, et al., 1999. Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nat Genet*, **22**, pp. 37–43.

Yan, L.X., Huang, X.F., Shao, Q., Huang, M.Y., Deng, L., Wu, Q.L., Zeng, Y.X. and Shao, J.Y., 2008. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA*, **14**(11), pp. 2348-60.

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T., et al. 2006. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, **9**, pp. 189–198.

Yoshihara, K., Tajima, A., Adachi, S., Quan, J., Sekine, M., Kase, H., Yahata, T., Inoue, I. and Tanaka, K., 2011. Germline copy number variations in *BRCA1*-associated ovarian cancer patients. *Genes Chromosomes Cancer*, **50**, pp. 167-177.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J. and Song, E., 2007. let-7 regulates self-renewal and tumorigenicity of breast cancer cells. *Cell*, **131**, pp. 1109–1123

Yuan, T.L. and Cantley, L.C., 2008. PI3K pathway alterations in cancer: variations on a theme. *Oncogene*, **27**, pp. 5497–5510.

Yuan, H., Chen, J., Liu, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T., Lin, B. and Xie, Y., 2015. Association of PIK3CA Mutation Status before and after Neoadjuvant Chemotherapy with Response to Chemotherapy in Women with Breast Cancer. *Clin Cancer Res.*, **21**(19), pp. 4365-72.

Zamore, P.D., et al., 2005. Ribo-gnome: the big world of small RNAs. *Science*, **309**, pp. 1519–1524.

Zanetti-Dallenbach, R.A., Schmid, S., Wight, E., Holzgreve, W., Ladewing, A., Hahn, S. and Zhong, X.Y., 2007. Levels of circulating cellfree serum DNA in benign and malignant breast lesions. *Int J Biol Markers*, **22**, pp. 95-99.

Zhang, W.B., Pan, Z.Q., Yang, Q.S., Zheng and X.M., 2013. Tumor suppressive miR-509-5p contributes to cell migration, proliferation and antiapoptosis in renal cell carcinoma. *Ir J Med Sci*, **182**, pp. 621-627.

Zhao, J.J., Liu, Z., Wang, L., et al., 2005. The oncogenic properties of mutant p110alpha and p110beta phosphatidylinositol 3-kinases in human mammary epithelial cells. *Proc Natl Acad Sci U S A*, **102**(51), pp. 18443–18448.

Zhao, H., Shen, J., Medico, L., Wang, D., Ambrosone, C.B. and Liu, S., 2010. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. *PLoS One*, **5**(10), pp. e13735.

Zhao, R., Wu, J., Jia, W., Gong, C., Yu, F., Ren, Z., Chen, K., He, J. and Su, F., 2011. Plasma miR-221 as a predictive biomarker for chemoresistance in breast cancer patients who previously received neoadjuvant chemotherapy. *Onkologie,* **34**(12), pp. 675-80.

Zhao, Y., Deng, C., Wang, J., Xiao, J., Gatalica, Z., Recker, R.R. and Xiao, G.G., 2011. Let-7 family miRNAs regulate estrogen receptor alpha signaling in estrogen receptor positive breast cancer. *Breast Cancer Res Treat,* **127**(1), pp. 69-80.

Zhao, X., Qureshi, F., Eastman, P.S., Manning, W.C., Alexander, C., et al., 2012. Preanalytical effects of blood sampling and handling in quantitative immunoassays for rheumatoid arthritis. *J Immunol Methods,* **378**, pp. 72–80.

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

Zhou, W., Muggerud, A.A., Vu, P., Due, E.U., Sorlie, T., Borresen-Dale, A.L., Warnberg, F. and Langerod, A., 2009. Full sequencing of *TP53* identifies identical mutations within *in situ* and invasive components in breast cancer suggesting clonal evolution. *Molecular oncology,* **3**(3), pp. 214-9.

Zhou, X., Marian, C., Makambi, K.H., Kosti, O., Kallakury, B.V., Loffredo, C.A. and Zheng, Y.L., 2012. MicroRNA-9 as potential biomarker for breast cancer local recurrence and tumor estrogen receptor status. *PLoS One,* **7**(6), pp. e39011.

Zhu, W., Qin, W., Atasoy, U. and Sauter, E.R., 2009. Circulating microRNAs in breast cancer and healthy subjects. *BMC Research Notes,* **2**, pp. 89.

Zhu, H. and Fan, G.C., 2011. Extracellular/circulating microRNAs and their potential role in cardiovascular disease. *Am J Cardiovasc Dis,* **1**(2), pp. 138-149.

Zhu, J., Zheng, Z., Wang, J., Sun, J., Wang, P., Cheng, X., et al., 2014. Different miRNA expression profiles between human breast cancer tumors and serum. *Front Genet,* **5**, pp. 149.

Zhuo, W., Zhang, Y., Xiang, Z., Cai, L. and Chen, Z., 2009. Polymorphisms of TP53 codon 72 with breast carcinoma risk: evidence from 12,226 cases and 10,782 controls. *Journal of Experimental & Clinical Cancer Research,* **28**, pp. 115.