Université de Montréal

Development of a multi-gene PCR assay for the prediction of the response to hormone therapy in breast cancer

Par :
Carolyn Nessim, MD

Programme Sciences Biomédicales
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de M.Sc. en Sciences biomédicales

13 décembre 2013

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Development of a multi-gene PCR assay for the prediction of the response to hormone therapy in breast cancer

Présenté par :
Carolyn Nessim, MD

A été évalué par un jury composé des personnes suivantes :

Louis Gaboury, MD, PhD
Isabelle Trop, MD
Sylvie Mader, Ph.D.
André Robidoux, MD

Résumé

Deux tiers des cancers du sein expriment des récepteurs hormonaux ostrogéniques (tumeur ER-positive) et la croissance de ces tumeurs est stimulée par l'estrogène. Des traitements adjuvant avec des anti-estrogènes, tel que le Tamoxifen et les Inhibiteurs de l'Aromatase peuvent améliorer la survie des patientes atteinte de cancer du sein. Toutefois la thérapie hormonale n'est pas efficace dans toutes les tumeurs mammaires ER-positives. Les tumeurs peuvent présenter avec une résistance intrinsèque ou acquise au Tamoxifen. Présentement, c'est impossible de prédire quelle patiente va bénéficier ou non du Tamoxifen.

Des études préliminaires du laboratoire de Dr. Mader, ont identifié le niveau d'expression de 20 gènes, qui peuvent prédire la réponse thérapeutique au Tamoxifen (survie sans récidive). Ces marqueurs, identifié en utilisant une analyse bioinformatique de bases de données publiques de profils d'expression des gènes, sont capables de discriminer quelles patientes vont mieux répondre au Tamoxifen.

Le but principal de cette étude est de développer un outil de PCR qui peut évaluer le niveau d'expression de ces 20 gènes prédictif et de tester cette signature de 20 gènes dans une étude rétrospective, en utilisant des tumeurs de cancer du sein en bloc de paraffine, de patients avec une histoire médicale connue. Cet outil aurait donc un impact direct dans la pratique clinique. Des traitements futiles pourraient être éviter et l'indentification de tumeurs ER+ avec peu de chance de répondre à un traitement anti-estrogène amélioré. En conséquence, de la recherche plus appropriée pour les tumeurs résistantes au Tamoxifen, pourront se faire.

Mots-clés : Récepteurs hormonaux, Cancer invasif du sein, Récepteurs ostrogénique, Facteur prédictif, Profil d'expression, PCR, outil prédictif

Abstract

Two thirds of breast cancers express the estrogen receptor (ER-positive tumours) and estrogens stimulate growth of these tumours. Adjuvant therapy with anti-estrogens such as Tamoxifen and Aromatase Inhibitors has been shown to increase survival in breast cancer patients. This treatment is, however, not successful in all ER-positive tumours. Tumours can present intrinsic or acquired resistance to Tamoxifen. However, it is currently impossible to predict which patient will benefit from Tamoxifen therapy and which will not.

Preliminary studies in Dr. Mader's lab have identified 20 genes whose expression levels in tumours are able to predict the response to Tamoxifen therapy (disease-free survival). These markers, identified using bioinformatics analysis of published gene expression datasets, were able to discriminate patients that would respond best to Tamoxifen from those that did not.

The overall purpose of this study is to develop a PCR kit to monitor expression levels of these 20 genes and to test this 20-gene signature in a retrospective study using paraffin-embedded breast cancer tissues of patients with a known medical history. This tool may thus have a direct impact on clinical practice through the development of markers of therapeutic success for treatment with Tamoxifen and possibly Aromatase Inhibitors. Futile treatments would be avoided thus preventing needless side effects, and improved identification of ER+ tumours with a low chance of success to anti-estrogen therapy. This will facilitate research into more appropriate treatments for hormone resistant tumours.

Keywords : Hormone receptors, Invasive breast cancer, Estrogen receptor, predictive factor, expression profile, PCR, predictive tool

Table of Contents

Introductory Pages

SECTION 3. **MATERIALS AND METHODS**

SECTION 4. **RESULTS**

SECTION 5. **DISCUSSION AND CONCLUSION**

List of Tables

List of Figures

List of Abbreviations

ABI – Applied Biosystems Inc.

AF – Activating Functions

AI – Aromatase Inhibitor

BMI – Body Mass Index

BRCA1 – Breast Cancer Susceptibility gene 1

BRCA2 - Breast Cancer Susceptibility gene 2

cDNA – Complimentary DNA

CHUM – Centre Hospitalier de l'Université de Montréal

CSR – Core Serum Response

CYP2D6 - Cytochrome P450 2D6

DNA - Deoxyribonucleic acid

EBCTCG – Early Breast Cancer Trialists Collaborative Group

EGFR – Epidermal Growth Factor Receptor

EMT – Epithelial-Mesenchymal Transition

ER – Estrogen Receptor

ER+ - Estrogen Receptor Positive

ER- - Estrogen Receptor Negative

ER-α – Estrogen Receptor Alpha

ER-β – Estrogen Receptor Beta

ERBB2 - Erythroblastic leukemia viral oncogene homolog 2

ERE – Estrogen response elements

ESR1 – Estrogen Receptor 1

FFPE – Formalin-fixed paraffin embedded

HER-2 - Human Epidermal Growth Factor Receptor 2

HR – Hazard Ratio

IRIC – Institut de Recherche d'Immunologie et de Cancérologie

LDA – Low density array

mRNA – Messenger RNA

NSABP – National Surgical Adjuvant Breast and Bowel Project

PR – Progesterone Receptor

PR+ - Progesterone Receptor Positive

PR- - Progesterone Receptor Negative

Q-PCR – Quantitative Polymerase chain reaction

RIN – RNA Integrity Number

RNA - Ribonucleic acid

RR – Relative Risk

RS – Recurrence Score

RT – Reverse Transcription

RT-PCR – Reverse Transcription Polymerase chain reaction

SERM – Selective Estrogen Receptor Modulator

SERD – Selective Estrogen Receptor Downregulator

TNBC – Triple Negative Breast Cancer

UV - Ultraviolet

I would like to dedicate this paper to my husband Matt, and my family for their support

Acknowledgments

I would like to thank my supervisor Sylvie Mader for her guidance and teachings. I would also like to thank my supervisor André Robidoux for his support and helping me initiate this project. I would like to acknowledge Slim Fourati for his significant contribution and help with this project, without whom this project would have not been possible. I would also like to thank the following members of the lab of Sylvie Mader for their support and teachings: Martine Bail, Edlie St-Hilaire, Maxime Parisotto, David Cotnoir-White, Laurence Fleury, Virginie Dupont and Marieke Rozendaal.

1. Introduction

1.1 Epidemiology

In Canada, one out of nine women will be diagnosed with breast cancer in their lifetime, by age 90[1]. Breast Cancer is the most common cancer in women and the second most common cause of cancer death in women. It is the main cause of death in women aged 40-59. Fifty percent of the cases can be explained by risk factors and 10% are found to have a positive family history. Only 5% of all breast cancers have known genetic mutations and syndromes such as BRCA1 and BRCA2. The majority of breast cancers are thus considered sporadic cancers[2].

In 2010, an estimated 23 200 women and 180 men in Canada were diagnosed with breast cancer. Approximately 445 Canadian women are diagnosed with breast cancer every week. In 2010, an estimated 5300 women and 50 men died of breast cancer in Canada. This means that, on average 100 women die of breast cancer in Canada, every week[1].

Globally, breast cancer incidence rates are highest in North America and Northern Europe and lowest in Asia and Africa. The incidence in China and Japan has been rising in recent years[3].

Since 1999, the incidence of breast cancer has remained quite stable, however, since 1986 the death rate from breast cancer has declined by more than 30%. This improvement in survival rate is most likely due to improvements in treatment strategies as well as better screening for breast cancer[1]. The decrease in mortality has been especially noted in women younger than 50[3] and women with ER/PR positive tumours[4]. At present, the five-year relative survival rate for female breast cancer is 87% (84% for men)[1].

1.2 Risk Factors

There are many risk factors that have been found to be associated with breast cancer. Some risk factors are stronger than others.

Gender and Age

Gender and age are among the strongest risk factors for breast cancer. Women are afflicted with breast cancer 100 times more frequently than men. In general, the older the person, the higher the risk. The incidence rises sharply with age until about the age of 45-50 and then the rise becomes less steep. At age 75-80, the incidence curve flattens and slightly decreases, as most women at this stage are menopausal and thus have less estrogen stimulation (Figure 1, p. 2)[5].

Age-specific SEER incidence, rates of female breast cancer per 100,000, 2000-2003

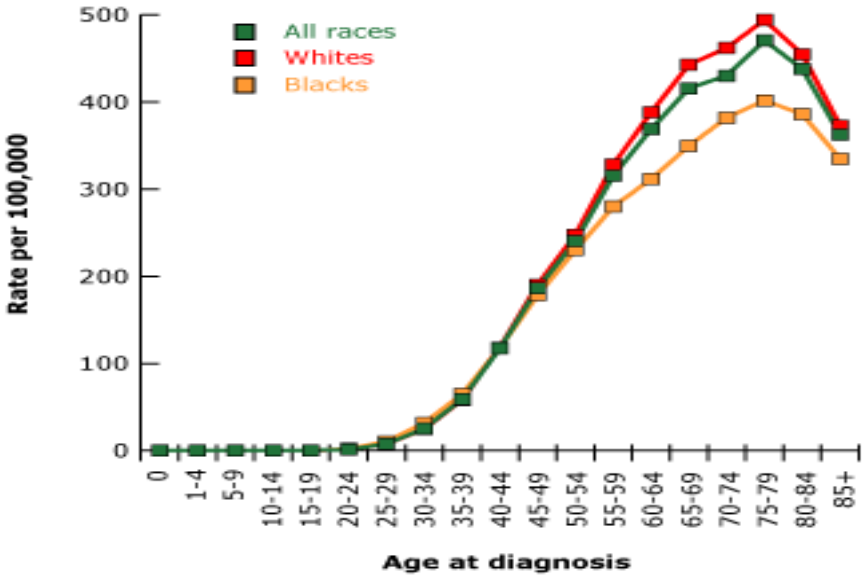


Figure 1: Incidence rate of breast cancer by age and race for 2000-2003

Race and Ethnicity

Breast cancer is more common in whites and less common in Hispanic and African American women. However, African American women tend to have more aggressive, hormone receptor negative cancers at a younger age and thus a lower survival rate[3].

Benign breast lesions

Proliferative benign diseases with cytological atypia increase the risk for breast cancer. Atypical lobular hyperplasia or atypical ductal hyperplasia, have a 4-6 fold relative risk (RR) of developing breast cancer and this becomes a 10-fold risk when the atypia is multifocal[6].

Personal History of Breast Cancer

A personal history of invasive breast cancer or a ductal carcinoma in situ also increases a person's risk of having a cancer in the contralateral breast. With in situ lesions, the 10-year risk of developing an invasive cancer in the contralateral breast is 5%. In patients that have already had an invasive cancer the risk of developing a contralateral breast cancer is 1% in premenopausal women and 0.5% in postmenopausal women[7].

Family history and genetic risk factors

A positive family history is an important risk factor, however, it is only reported by 15-20% of women diagnosed with breast cancer. In a pooled analysis done in 2001 by the Collaborative Group on Hormonal Factors in Breast Cancer, data was used from over 50 000 women with breast cancer and 100 000 controls. The results showed that the risk of breast cancer for a woman with one affected first-degree relative was increased 1.80 fold. With two affected first-degree relatives, the risk is increased 2.93 fold. The risk ratios were highest for women with young affected relatives. Thus, the risk was increased 2.9 fold for a woman

whose relative was diagnosed before age 30, but only 1.5 fold increased if the affected relative was diagnosed after age 60. Similarly, if one relative had breast cancer before age 40, the risk of breast cancer was increased 5.7-fold[8].

Specific genetic mutations that predispose to breast cancer are very rare; only 5 to 6% of all breast cancers are directly attributable to inheritance of a breast cancer susceptibility gene (Table I, p. 5). Genetic mutations in these genes are often associated with various cancer syndromes, where patients may be afflicted with more than one type of cancer and various diseases. These germ-line mutations, which are often associated with triple-negative breast cancer (TNBC), are beyond the scope of this thesis, which focuses on the prognostic role of somatic mutations.

Table I: Germ Line Mutations in Breast Cancer

Mutation	Associated Cancers/Diseases
BRCA1	• Breast Cancer • Ovarian Cancer (higher risk than BRCA2) • Cervical Cancer • Uterine Cancer • Pancreatic Cancer • Colon Cancer • Male Breast Cancer (lower risk than BRCA2) • Testicular Cancer • Prostate Cancer
BRCA2	• Breast Cancer • Ovarian Cancer • Pancreatic Cancer • Gastric Cancer • Gall bladder Cancer • Bile duct Cancer • Melanoma • Male Breast Cancer (more common BRCA2) • Prostate Cancer (more common BRCA2)
ATM	• Breast Cancer • Ataxia-telengectasia disease
p53 (Li Fraumeni Syndrome)	• Breast Cancer • Soft tissue and Bone Sarcoma • Leukemia • Brain tumours
CHEK2 (Li Fraumeni Sydrome)	• Same as p53
PTEN (Cowden Syndrome)	• Breast Cancer • Benign breast diseases • Digestive tract tumours • Thyroid tumours • Uterine tumours • Ovarian tumours
CDH1	• Breast Cancer (Invasive lobular carcinoma) • Gastric Cancer
STK11/LKB1 (Peutz-Jeghers Syndrome)	• Harmartamous polyps in GI tract • Pigmented macules on lips, buccal mucosa, digits • Digestive Cancers • Cervical Cancer • Lung Cancer • Testicular Cancer • Ovarian Cancer • Uterine Cancer

Reproductive and Hormonal Risk Factors

Prolonged exposure to endogenous estrogen has also been shown as an important risk factor for breast cancer. For every two-year delay in the onset of menarche, there is a 10% reduction in cancer risk[8]. Moreover, the risk increases as menopause is delayed. The RR increases by 1.03% for each year older at menopause[9]. This was further emphasized when it was noted that women that had had a bilateral oophorectomy before the age of 40 had a 50% lifetime risk reduction of breast cancer[10].

Data from the Nurses' Health Study suggest that the association is strongest for hormone receptor-positive breast cancers. Endogenous hormone levels were measured in 322 women who developed breast cancer and in 643 age-matched controls without breast cancer. When the highest and lowest quartiles of serum hormone concentration were compared, there was a significant direct association between breast cancer risk and levels of both estrogens and androgens. However, the association was strongest when the analysis was restricted to ER and PR-positive tumours, and in situ tumours[11].

Breast density also seems to have an association with breast cancer risk. The denser breasts (greater than 75% density) compared to women of the same age with less or no dense tissues have five times greater risk of developing breast cancer[12]. Both endogenous and exogenous estrogen may influence mammographic density. Mammographic density decreases after menopause when ovarian function declines. Hormonal replacement therapy, with combination of estrogen and progesterone, increases mammographic density [13], while tamoxifen, which has antiestrogenic effect, decreases mammographic density [14]. Mammographic density therefore can be regarded as a marker of the effect of estrogen on the

breast tissue. To what extent mammographic density is a predictor for both hormone receptor-positive and hormone receptor-negative tumors is still unclear.

Pregnancy related factors

Nulliparous women are at increased risk for breast cancer (RR 1.2-1.7). Moreover, the younger the woman at her first time full-term pregnancy the lower the risk. The risk is 20% lower if the first birth is at age 20, 10% lower for a first birth at age 25 and 5% higher if the first birth is at age 35[9].

Exogenous hormone factors

It is controversial whether or not long-term use of oral contraceptives increases the risk of breast cancer and data are conflicting. Long-term hormone therapy replacement with estrogen and progesterone has however been shown to increase the risk of breast cancer, especially hormone positive cancers. It must be noted, however, that women taking an unopposed estrogen therapy have a slightly lower risk[15, 16].

Ionizing radiation

Exposure to ionizing radiation has been shown to greatly increase the risk of breast cancer. Patients who have received radiotherapy treatment for Hodgkin's lymphoma to the chest wall, especially between the ages of 10-16 and up to the age of 45, are at increased risk of getting breast cancer[17].

Lifestyle and Dietary Factors

Women of higher socioeconomic status are at a two-fold greater risk for breast cancer. This is thought to be due to differing educational, occupational and economic level reflecting

different reproductive patterns with respect to parity, age at first birth, age at menarche and utilization of screening mammography.

In postmenopausal women, it has been shown that a higher body mass index (BMI) is associated with a higher risk of breast cancer. Obese postmenopausal women have higher estrogen levels than non-obese postmenopausal women, due to the conversion of adrenal androgens to estrogens in fatty tissue. In a pooled analysis of seven prospective studies in the US, women who weighed at least 80 kg (176 lbs., BMI >33 kg/m^2) had a 25% higher risk of breast cancer as compared to those weighing less than 60 kg (132 lbs., BMI<21 kg/m^2), after adjusting for height[18]. In the same-pooled analysis, the opposite association was found in premenopausal women. Those with a BMI≥31 kg/m^2 were 46% less likely to have breast cancer compared to those women who's BMI was <21 kg/m^2 [18].

Alcohol Intake

Many dietary risk factors have been evaluated but are quite difficult to interpret with regards to a direct causal relationship to breast cancer. Increased alcohol intake is the only dietary risk factor that has been consistently shown in several epidemiological studies, to increase the risk of breast cancer. More specifically, it increases the risk of hormone positive breast cancers and the use of hormone replacement therapy acts as an additive risk factor to increased alcohol intake. It is believed that it may be in part due to the increased estrogen and androgen levels in women who consume alcohol as well as increased mammary gland susceptibility to carcinogenesis and DNA damage in women who consume alcohol[19]. Moderate to increased use (≥ three drinks per day) as compared to those who abstain from drinking, has been shown to have a 12% increased risk of breast cancer[20].

In conclusion, most of these risk factors for breast cancer are associated with increased exposure to estrogen, whether it be endogeneous or exogeneous. The higher or the longer breast tissue is exposed to estrogen, the higher the risk of breast cancer.

1.3 Molecular Subtypes of Breast Cancer

Breast cancer is a heterogeneous and phenotypically diverse disease. Classically, pathologic features of the cancer have been used to determine prognosis. These features include the histologic grade of the tumour, the presence of lymphovascular invasion, the presence of nodal disease as well the expression of various receptors, namely; ER, PR and Her-2-neu. These features have helped sub-categorize breast cancers into different groups with different tretment options.

However, more recently, due to the progress in molecular profiling and using gene expression arrays, Perou *et al*. further classified these 3 subtypes of breast cancer (ER/PR, Her-2 and TNBC) at a genetic level, and characterized other biologic subtypes[21],[22],[23]. As these different subtypes have distinct responses to therapy, this molecular portrait of breast cancer has further helped in determining prognosis, and may eventually aid in treatment strategies.

Five different intrinsic subtypes of breast cancer have been identified, each having their own distinct genetic profile.

Luminal Subtype

The most common subtype is the Luminal subtype which is further subdivided into luminal A and luminal B; making up two distinct intrinsic subtypes. They make up the majority of ER+ breast cancers. These tumours typically express luminal cytokeratin 8 and 18

and are characterized by their expression of ER, PR and other genes associated with ER activation. The subdivision is not only at a molecular level but also corresponds to different clinical outcomes prognostically.

Luminal A tumours, making up approximately 40% of all breast cancers, usually have a high expression of ER-related genes, low expression of HER2 genes and low expression of proliferation-related genes. As expected, they correspond to the best prognosis.

Luminal B tumours, making up approximately 20% of all breast cancers, have a lower to moderate expression of ER-related genes, variable expression of HER2 and a higher expression of proliferation genes. They relapse more frequently on antiestorgen/aromatase inhibitor therapy and thus have a worse prognosis than luminal A cancers.

HER-2 enriched Subtype

The second most common subtype is the ERBB2-positive or HER2-enriched subtype. These make up approximately 10-15% of all breast cancers and are characterized by high expression of the HER2 and proliferation gene clusters and a low expression of the luminal cluster. They are typically ER-PR-negative. Although this subtype has a poorer prognosis, with the advent of targeted therapy against HER2, Herceptin, the outcome of patients has greatly improved[24].

Basal-like Subtype

The third subtype has been named the Basal-like subtype because of its similarity in expression to that of basal epithelial cells. They make up approximately 15-20% of all breast cancers. They have low expression of luminal and HER2 cluster genes. These tumours are usually ER/PR-negative and HER2-negative. Naming these breast cancers however, “triple-

negative" is a misnomer. Although most "triple-negative" tumours are basal-like and most basal-like tumours are triple-negative, there is a significant 30% discordance between the two types. Basal-like tumours have a high expression of the proliferation cluster of genes and are almost always histologically high-grade tumours. They demonstrate widespread genomic instability and have a high expression of the epidermal growth factor, and basal epithelial cytokeratin 5, 14 and 17. Eighty percent of BRCA1 mutation carriers have basal-like tumours[25]. These tumours have a poor prognosis, as they do not benefit from established targeted therapies, being mostly receptor negative. However, they do respond to chemotherapy, with a complete pathologic response rate of up to 45%, which is promising.

Claudin-low Subtype

The fourth subtype is the non-basal TNBC. This subtype is more uncommon however clinically quite significant. They have an extremely low to absent expression of the luminal cluster genes and high expression of the epithelial-mesenchymal transition (EMT) genes, immune response genes and characteristics reminiscent of stem cells. Many studies have been performed regarding the EMT process and suggest its implication in tumour progression and spread of metastasis[26], which may account for the poor prognosis of TNBCs. These tumours tend to respond to chemotherapy at an intermediate level between the basal-like and luminal tumours.

Normal-like Subtype

This final subtype is the hardest to characterize clinically, however is always present in gene expression arrays. It is difficult to know whether or not this is a true subtype or a technical artefact, due to low tumour cell composition of those specifically sampled

specimens. It has a similar gene expression pattern as normal breast tissue.

Understanding these different intrinsic subtypes has greatly aided in the understanding of the biology of these tumours, which can lead to better evaluation of treatment strategies for breast cancer.

1.4 ER/PR-positive Breast Cancer

Hormone receptor positive breast cancers are the most common type, making up approximately two thirds of all breast cancers. Hormone receptor-positive breast cancer was characterized over 40 years ago by Elwood Jensen. He noted that radiolabeled estrogens concentrated preferentially in some human breast tumors as well as in estrogen target organs. These findings then led to the discovery of the estrogen receptor (ER), as well as the progesterone receptor (PR), which were found in high abundance in a large fraction of malignant breast tumours. It has since become clear that human breast cancers are dependent upon estrogen and/or progesterone for growth and that this effect is mediated through ERs and PRs. As mentioned previously, these tumours are now classified as the luminal subtype of breast cancers. It is these luminal tumours that will be the focus of this thesis.

Molecular biology and physiology of the estrogen receptor

Estrogens have multiple actions on various sites including the cardiovascular, skeletal, immune, gastrointestinal and neural systems; however, their most important action is on the reproductive organs[27]. They reprogram gene expression via the activation of nuclear estrogen receptors. These receptors bind to estrogen with high affinity and specificity and function as ligand-modulated nuclear transcription factors[28,29].

Two ER molecules have been identified, namely ER-alpha (ER-α) and ER-beta (ER-β)[30]. The key functional domains in these receptors are the C or DNA-binding domain, which binds with high affinity and specificity to specific DNA sequences (estrogen response elements (ERE)) in the promoter regions of target genes, and the E or ligand-binding domain, which bind estrogens and estrogen analogues[31]. The ERE in the target genes is a 15-base pair inverted-repeat DNA sequence (RGGTCAnnnTGACCY), to which ER dimers bind with high affinity and specificity[32], with one receptor molecule in contact with each five base-pair segment of the response element[33]. The estrogen receptors contain two regions, termed activation functions (AF) that mediate the increase in transcriptional activity induced by the receptors in the presence of ligand. AF-1, located near the amino-terminal end of the receptor, acts independent of ligand, whereas AF-2, located in the ligand-binding domain, is ligand dependent[34]. There are also numerous co-regulator molecules, including RNA cofactors that interact with the receptors in a ligand-dependent manner, modulating receptor-mediated transcription by interacting with both AF regions and transcription factors associated with RNA polymerase II[35].

When an estrogen or its analogue reaches the cell nucleus and binds to ER, the conformation of the ligand-binding domain of the receptor changes, either allowing or preventing interaction with the co-activators, depending on whether the ligand is an agonist or an antagonist, respectively. The estrogen receptor dimers bind to the ERE in target genes, and via agonist-dependent association with co-activators, increases the rate of transcription by interacting with and activating necessary components of the transcriptional apparatus. Moreover, the ability of steroid hormone receptors to activate transcription of endogenous genes likely depends upon their ability to affect chromatin structure. Many steroid hormone

receptors interact with coregulator proteins that are implicated in the remodeling of local chromatin structure and the acetylation of histones[36,37]. In fact, enhancement of transcription by adding ligand to ER was observed using chromatinized template DNA but not when using naked DNA lacking histones[38]. This is referred to as the genomic classical mode of action (Figure 2, p. 15).

In the genomic non-classical mode of action, ER regulates gene expression without interacting with DNA directly. It acts via other transcription factors such as Fos/Jun activating protein-1 (AP-1) complex (Figure 2, p.15)[39].

ERs can also function independently of estrogen. Both epidermal growth factor and insulin-like growth factor-1, acting via their extracellular membrane bound receptors, can stimulate transcription of ER target genes in the absence of estrogen[40,41]. Therefore, cross-talk and signal amplification occurs between growth factor signaling pathways and nuclear receptors[42].

Estrogens also have non-genomic actions. They bind with high affinity to other cell components, including plasma membranes. Some effects of estrogen, such as rapid induction of MAP kinase and Erk pathways, appear to involve direct action of estrogen receptors at the plasma membrane rather than genomic modulation (Figure 2, p.15). As these rapid effects occur without ER-gene interaction, they are called "non-genomic," although the signals initiated by these mechanisms ultimately result in regulation of genes. These responses are observed in diverse tissues, including the cardiovascular system, central nervous system, and in breast cancer cells.

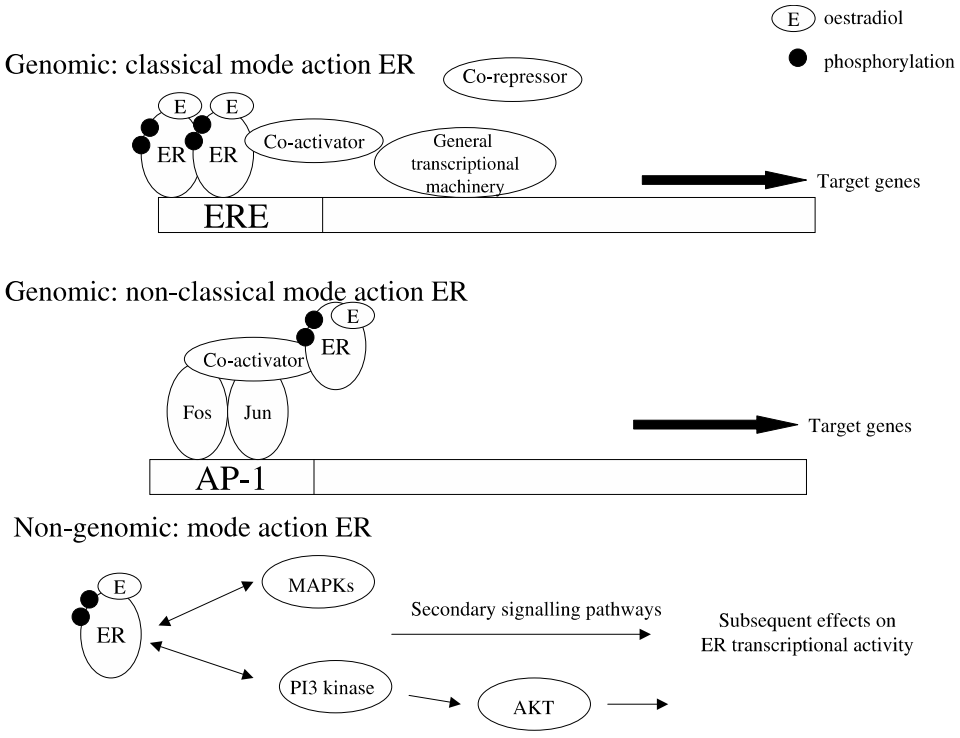


Figure 2: The different modes of action of ER (Genomic and Non-genomic)

Estrogens and the mammary gland

As female mice mature, the rudimentary ductal tree of the mammary gland elongates in response to estrogens and branches in response to progesterone to fill the stroma. In ER-α deficient mice, the ducts fail to elongate[43]. If ER-β is deficient the gland develops normally and the mice can nurse their young with a normal lactation function. If both ER-α and ER-β are deficient, the phenotype is similar to those mice with ER-α deficiency only, emphasizing the importance of ER-α in male and female reproduction[44].

1.5 Hormone Therapy for ER/PR positive breast cancer

Currently there are three main anti-estrogen therapies that can be used in the treatment of ER+ breast cancer: Tamoxifen, Aromatase Inhibitors (AI) and Fulvestrant. Clinically,

Tamoxifen and AIs are the treatments most often used, while fulvestrant is mainly used as second line therapy in patients with metastatic disease that have not responded or are progressing on standard therapy with Tamoxifen or AIs.

Tamoxifen

Tamoxifen is a competitive inhibitor of estrogen binding to ERs, and has a mixed agonist and antagonist activity, depending on the target tissue. It is therefore called a selective estrogen receptor modulator (SERM).

Its physiological effects in postmenopausal women can illustrate this mixed antagonist/agonist effect. Tamoxifen is known to provide some protection against menopausal bone loss[45] and lowers serum total and low-density lipoprotein-cholesterol concentrations[46,47] via its agonistic properties. This has been shown to lead to less osteoporotic fractures. Whether it is protective against cardiovascular disease is still controversial. Other agonistic features include the induction of endometrial hyperplasia and the increased risk of endometrial cancer. Moreover, Tamoxifen increases the risk of thromboembolic events such as deep vein thrombosis and pulmonary embolism as well as increases the risk of stroke and cataracts[47].

The most salient feature, however, of Tamoxifen is its antagonistic properties with respect to breast cancer. Among women with ER+ breast cancer, Tamoxifen reduces the risk of recurrence and death and prevents the development of contralateral breast cancer when given as adjuvant therapy for early stage disease and can provide palliation in patients with metastatic disease[47,48,49]. Other side effects due to is antagonistic properties include hot flashes and vaginal dryness.

The molecular mechanisms by which SERMs can act as ER agonists in one tissue and as antagonists are just starting to be better understood. The change in receptor conformation that follows the binding of Tamoxifen to the ER modulates interactions with co-repressors and co-activators that are required for ER-mediated gene regulation. Co-activators increase transcriptional activity by promoting the interaction between the receptor and the transcriptional apparatus and chromatin remodeling[50], whereas co-repressors restrain ER activity, maintaining the receptor in protein/DNA complexes that do not promote transcription and favor an inactive conformation of chromatin[51]. Thus, the main mechanism of action of Tamoxifen is the induction of an abnormal receptor conformation and altered recruitment of cofactors. The complement of co-activators/co-repressors expressed in different tissues may therefore dictate whether the receptor is active or inactive.

Aromatase Inhibitors

The aromatase inhibitors decrease circulating levels of estrogen in postmenopausal women by blocking the action of the enzyme, aromatase, which converts androgens to estrogens in peripheral tissues[52]. There are three different AIs used in the clinic, namely, Anastrazole, Letrozole and Exemastane. These agents are used in the treatment of postmenopausal patients with breast cancer in the adjuvant and metastatic setting. AIs are contraindicated in premenopausal women, since they may increase estrogen levels via a reduced feedback of estrogen to the hypothalamus and pituitary. This would lead to an increase in gonadotropin secretion and stimulation of the ovary, and ultimately to an increased concentration of the precursors of estrogens, androgen and increased expression of aromatase[53]. Contrary to Tamoxifen, AI's are not associated with an increased risk of thromboembolic events or endometrial cancer, however they do have a similar profile with

regards to hot flashes and vaginal dryness and have additional musculoskeletal side effects including osteoporosis and arthralgia.

Fulvestrant

Fulvestrant is a selective estrogen receptor down-regulator (SERD). It has a steroidal structure that binds competitively to the estrogen receptor (ER), with high affinity, and downregulates ER by functional blockade and increased turnover. After binding to fulvestrant, degradation of the ER is accelerated, ultimately resulting in a reduction in cellular ER. The downregulation of cellular levels of the ER protein results in complete abrogation of estrogen-sensitive gene transcription. As a pure estrogen antagonist, fulvestrant avoids the risk of detrimental side effects of selective ER modulators such as tamoxifen, which has partial agonist activity. Due to its unique mode of action, fulvestrant lacks cross-resistance with existing agents. Fulvestrant, an antiestrogen classified as an estrogen receptor antagonist without known agonist effects is mainly used for the treatment of postmenopausal, hormone receptor-positive women with progressive metastatic breast cancer after antiestrogen therapy. The main adverse effects associated with therapy are nausea, asthenia, pain, vasodilation and headache[54,55,56].

Efficacy of Hormone Therapy

Multiple randomized clinical trials have been performed to evaluate the efficacy of hormone therapy in breast cancer. One of the first trials to demonstrate a benefit for Tamoxifen therapy was the NSABP B-14 trial, which compared 5 years of adjuvant Tamoxifen to placebo. Since that trial multiple others were performed. In 2011, the Early Breast Cancer Trialists Collaborative Group (EBCTCG) performed a meta-analysis of

randomized trials, which compared tamoxifen versus no endocrine treatment in premenopausal and postmenopausal women[49]. With a median follow-up of 13 years, tamoxifen resulted in a reduction in breast cancer recurrence by 39% compared to placebo, which translated into a 15-year absolute reduction of 13% (33% versus 46%). This was seen in patients with both node-negative and node-positive ER+ breast cancer. There was no effect on recurrence for patients with ER-negative breast cancer. It also showed a reduction in risk of breast cancer mortality by 30%, which translated into a 15-year absolute reduction of 9% (24% versus 33%). The magnitude of benefit was similar in women less than 45 years of age and in women between the ages of 55 and 69 years.

With regards to AI's, multiple trials have also been performed. In 2010, the EBCTCG performed a meta-analysis of these trials demonstrating the benefit of AI's compared to Tamoxifen[57]. This meta-analysis showed that, with a mean follow-up of 6 years, treatment with an AI as a single agent therapy had a reduction in the risk of recurrence compared to Tamoxifen (3% absolute reduction in 5-year risk of recurrence, 12% versus 15%). There was no difference between an AI and Tamoxifen with regards to overall survival. A secondary analysis, with a mean follow-up of 4 years, evaluated the use of Tamoxifen for 2-3 years and then switching to an AI for the last 2-3years to complete 5 years versus staying on Tamoxifen for the entire 5 years of treatment. This showed that switching to an AI reduced the risk of recurrence by 3% and reduced 5-year breast cancer mortality by 2%. Finally, with regards to sequencing therapy, in the National Cancer Institute of Canada Clinical Trials Group MA17 study, 5 years of AI versus placebo was given to patients who had completed 5 years of Tamoxifen. With a median follow-up of 64 months, treatment with letrozole improved disease free and overall survival compared to placebo[58].

As already mentioned, Fulvestrant has been shown to be an effective and well tolerated treatment for patients with metastatic breast cancer when compared to Tamoxifen and AI's and is especially useful in patients with Tamoxifen resistance[55]. It, however, has not been shown to be superior to Tamoxifen or AIs and is much more costly, explaining its limited use clinically. Futher studies are necessary to evaluate its efficacy as adjuvant therapy.

Thus, in summary patients with ER/PR positive breast cancer greatly benefit from hormone therapy after surgery and in the metastatic setting. Unfortunately, however, despite the benefits of these therapies, 40% of patients still recur and eventually succumb to their disease. To date, we have an inability to identify which patients will respond and which will not. Multiple studies have been performed to understand the resistance mechanisms involved with regards to hormone therapy, most of these studying the resistance to Tamoxifen.

1.6 Resistance to Tamoxifen

Research over the last two decades has identified two forms of resistance to Tamoxifen therapy: Intrinsic (de novo) resistance, in which ER-negative and many ER+ tumours do not respond to Tamoxifen at the outset of therapy, and acquired resistance, where ER+ tumours that initially responded to therapy stop responding and may actually exploit the Tamoxifen-ER complex as a stimulator as opposed to an inhibitory signal[59]. As this is quite a complex and exhaustive topic this will simply be summarized here.

Intrinsic Resistance

Understandably, it has been noted that ER/PR-negative breast cancers do not respond to Tamoxifen therapy. However, it has also been noted that approximately 25% of ER+/PR+ tumours, 66% of ER+/PR- and 50% of ER-/PR+ tumours fail to respond to or develop early

resistance to Tamoxifen[60]. A number of factors have been identified that may contribute to the intrinsic resistance of Tamoxifen.

Loss of ER-α expression/function

Lack or loss of ER expression could confer resistance. This is the dominant mechanism of intrinsic resistance to Tamoxifen, with the majority of ER-/PR- breast cancers not responding to Tamoxifen or AIs. Although quite rare (<1%), mutations in coding of the ER gene alter the effects of bound anti-estrogens, leading to a hypersensitive receptor, with enhanced binding of co-activators in the presence of low estrogen levels. These somatic mutations alters the crosstalk between ER-*α* and various ER-*α* pathways that normally down-regulate ER signalling. Such loss of regulation could theoretically enhance ER-mediated cell growth and contribute to the development of resistance. Also, epigenetic changes have been identified that cause transcriptional inactivation of the ER gene[39].

Altered expression of ER-β

Although the role of ER-β in Tamoxifen resistance remains unclear, it has been shown that relative changes in the expression of the ER isoforms that occur during tumorigenesis parallel the marked changes in estrogen action. Interestingly, in an RT-PCR study, the median ER-β mRNA levels were approximately 2-fold higher than ER-α levels in tamoxifen-resistant tumours compared with tamoxifen-sensitive tumours[39].

Tissue-specific availability of co-activators and co-repressors

As already mentioned, when Tamoxifen is bound to an ER it changes the conformation of the ligand-binding domain, generating an abnormal receptor conformation, recruiting co-repressors, and thus leaving ER in an inactive state. In some cells, tamoxifen-induced AF2

inhibition may be bypassed when enough co-activator function is recruited to the ligand-independent domain, AF1[61]. In other cell types, available co-activator proteins might bind to and activate AF2 despite the presence of tamoxifen[62].

Modulation of ER expression through second messengers

As discussed before, ER can be activated independent of estrogen via growth factor signalling. Both ER expression and function correlate with distinct patterns of growth factor receptor overexpression. It appears likely that an appropriate growth factor environment is necessary for efficient mitogenesis in breast cancer cells, with steroid hormone and growth factor signalling pathways "cross talking" to reinforce each others' signalling. One proposed model for both primary and secondary hormone resistance in breast cancer is that phenotypic changes in growth factor signalling pathways may perturb this balance of steroid hormone and growth factor interaction, providing a selective advantage for tumour cell proliferation[63], potentially explaining the resistance to endocrine therapy in breast cancer.

As an example, ER- and ER+ but PR- tumours overexpress proteins of the epidermal growth factor receptor (EGFR) family, particularly EGFR and the HER2 protein[63]. Studies have shown that ER expression is suppressed when HER2 or EGFR receptor is activated, leading to resistance to Tamoxifen[64]. Others suggest that the antagonist activity of tamoxifen on the ER may be diminished via an interaction between HER2 and AIB1, an ER co-activator[65].

As another example, when ER is activated by tyrosine kinase receptors in response to growth factor stimulation, PI3K (phosphatidyl-inositol-3-OH kinase) catalyses the formation of PIP3. One of the downstream targets of this pathway is AKT, whose activation promotes

cellular proliferation and anti-apoptotic responses. There is evidence that ERα can bind in a ligand-dependent manner with a regulatory subunit of PI3K, leading to the activation of AKT and subsequent downstream effects. However the relationship with ER is reciprocal, in that PI3K activates AKT, which phosphorylates the ER at serine-167 resulting in ligand-independent activation. Interestingly, *in vitro*, elevated levels of AKT confer Tamoxifen resistance[39].

Finally, in the presence of Tamoxifen, ER may interact with the stress-activated protein kinase/c-junNH2 terminal kinase pathway (SAPK/JNK) by binding with the AP-1 transcription complex. Tamoxifen-resistant tumours, compared with estrogen-treated tumours, have increased AP-1 dependent transcription and phosphorylated c-Jun and JNK levels. In addition, the conversion to a resistant phenotype has been associated with an increase in oxidative stress (as measured by increases in superoxide dismutases and glutathione-S-transferase). It has been shown that tamoxifen resistant tumours have high AP-1 DNA binding. This is due to the fact that tamoxifen can induce intracellular oxidative stress, which leads to activation of JNK and SAPK, which in turn increase the transcriptional activity of AP-1. This chain of events may explain the potentiation of the agonistic effects of tamoxifen at AP-1 sites in resistant tumours[39].

Modulation of ER-α expression by BRCA1

Tamoxifen resistance in patients with the BRCA1 mutation may be due to the fact that most of these patients are ER-negative. BRCA1-mutant tumors fail to express ERα due to the loss of BRCA1-mediated transcriptional activation of ESR1. Loss of the wild-type BRCA1 allele, which occurs during neoplastic development in BRCA1 mutation carriers, has a direct effect on ESR1 transactivation, resulting in the loss of ERα mRNA and protein expression[66].

Altered Tamoxifen metabolism

Tamoxifen is converted to its active metabolites, endoxifen and 4-hydroxytamoxifen, by two rate-limiting enzymes, cytochrome P450 2D6 (CYP2D6) and UDP-glucuronyltransferase-2B7 (UGT2B7)[67],[68]. Although it was initially thought that CYP2D6 polymorphisms may confer a relative resistance to Tamoxifen, multiple sub-analyses of several clinical trials (IBIS-1, NCCTG, BIG 1-98, ATAC)[69],[70],[71],[72] have not shown a difference in outcomes with regards to survival or recurrence for patients that were poor versus good metabolizers of the drug.

Acquired Resistance

Loss of ER-α expression/function

Approximately 20-30% of patients, that initially have ER+ tumours, treated with tamoxifen, acquire a resistance via loss of ER-α in the recurrent tumours. It is however important to note that even in those patients that relapse under Tamoxifen treatment, 20% of them will still respond to an AI or to the full antiestrogen Fulvestrant, suggesting that ER continues to regulate tumour growth even in tamoxifen-resistant patients[39].

Co-repressor and co-activator expression levels

Co-repressor and co-activator expression levels may influence the development of secondary resistance to tamoxifen. In animal models, prolonged tamoxifen exposure alters the balance between co-activators and co-repressors in favour of the agonist, growth-promoting properties of tamoxifen; the net effect is stimulation of growth despite the continued presence of tamoxifen[73]. This is accompanied by suppression of co-repressor N-CoR levels in the tamoxifen-stimulated tumours when compared with their tamoxifen-sensitive counterparts[74].

Growth factor pathways

As already mentioned in *de novo* tamoxifen resistance, growth factor pathway "cross talk" also plays a role in acquired tamoxifen resistance. Signalling through EGFR and the HER2 receptor appears to bypass the estrogen requirement for breast cancer cell growth and may drive initially ER+ cells into an endocrine therapy-resistant state[65,75]. It is postulated that activation of growth factor pathways such as these modulates ER activity via phosphorylation, which alters its function, especially its ability to interact with tamoxifen[76,77]. The net result is that an ER+ cell becomes "hormone-independent" and therefore resistant to tamoxifen.

1.7 DNA Microarray versus RT-QPCR

Important goals of cancer research include the discovery of novel cellular targets to exploit for novel targeted treatments, new biomarkers for early cancer detection, and to provide a better classification of cancers for prognostication and treatment selection.

Toward this end, a significant effort has been devoted to understanding the molecular basis of carcinogenesis and the biologic behavior of human cancers. Carcinogenesis is a multistep process involving genetic and epigenetic events that result in altered expression of numerous genes[78]. Confounding this complexity, many of the so-called oncogenes and tumor suppressor genes are signaling molecules, which control the expression of a subset of downstream genes. Cells respond to environmental signals by modulating the expression of genes contained within the nucleus. When genes are activated, they are transcribed to generate messenger RNA, which is transported from the nucleus to the cytoplasm and translated into protein by the ribosomes[79].

Approximately 3 to 5 percent of genes are active in a particular cell, even though all cells have the same information contained in their DNA. Most of the genome is selectively

repressed, a property that is governed by the regulation of gene expression, mostly at the level of transcription (ie, the production of messenger RNA from the DNA). In response to a cellular perturbation, changes in gene expression take place that result in the expression of hundreds of gene products and the suppression of others. This molecular heterogeneity is thought to underlie, at least in part, the variability in outcome and response to therapy that characterizes tumors of different histology. Significant variability also exists for tumors of a specific histologic type. In general, clinical management decisions and prognostic estimates are based solely upon histopathologic analysis of tumor tissue. However, tumor behavior cannot be adequately understood through the analysis of one or a small numbers of genes, particularly for the common solid tumors[79].

DNA Microarray

The examination of multiple expressed genes and/or proteins provides more useful information for both classification and prognostication of individual tumors. The development of microarray methodology, which permits the expression of thousands of genes to be assayed simultaneously, represents a powerful technique to read the "molecular signature" of an individual patient's tumor. This process is termed gene expression profiling. Analyzing gene expression patterns across individual patients with the "same" disease may reveal molecular differences. Such classification may allow better treatment selection and prognostication.

The biggest advantage microarray technology has to offer is the large number of transcripts that can be quantified in a single experiment. DNA microarrays are capable of making tens of thousands of gene expression measurements simultaneously. Major commercial suppliers of DNA microarrays have recently released products in which the entire complement of known expressed human genes (the “transcriptome”; approximately 40,000

expressed sequences) can be measured on a single microarray. The unprecedented ability to monitor the expression of entire genomes has led to biological discoveries that would not have been possible by other methods. Nevertheless, microarray technology has limitations including its relatively high cost and inability to analyze more than one sample per array experiment[80]. Moreover, analysis of data is quite challenging and based on calculating the ratio of signal intensity between tissues based on signaling from fluorescent detectors (eg. tumour vs. normal, treated vs. untreated). This tool is best used for the discovery of candidate genes, as it analyses thousands of genes at once. Once these genes are discovered however, this smaller group of genes is now best suited to be studied using Real-time Reverse Transcription-Quantitative Polymerase Chain Reaction (RT-QPCR), as this is better suited to analyzing multiple samples at once.

RT-QPCR

Real time RT-QPCR is the gold standard by which other methods are compared. This technology not only provides a tissue's genetic profile but it does so in a very quantitative method, requiring very small amounts of cDNA. RT-QPCR measures the accumulation of PCR product, with each PCR cycle. The main advantage of this method is its relative simplicity of experiments as well as the ability to obtain a quantitative result in a single reaction. Moreover, hundreds of samples (or genes) can be analysed simultaneously. The analytical precision of QPCR is superior to other methods of genetic profiling. Very small amounts of cDNA are required to run experiments which is especially useful when extracting RNA from formalin-fixed paraffin embedded (FFPE) tissues which often provides a low yield of degraded RNA.

Interestingly, the capabilities of microarray technologies and RT-QPCR are starting to overlap with companies offering high-density microarrays that are designed for the analysis of relatively smaller numbers of genes and for high throughput analysis platforms allowing the analysis of multiple samples at once. Conversely, companies are making advances in RT-PCR technology enabling simultaneous analysis of larger numbers of genes or samples.

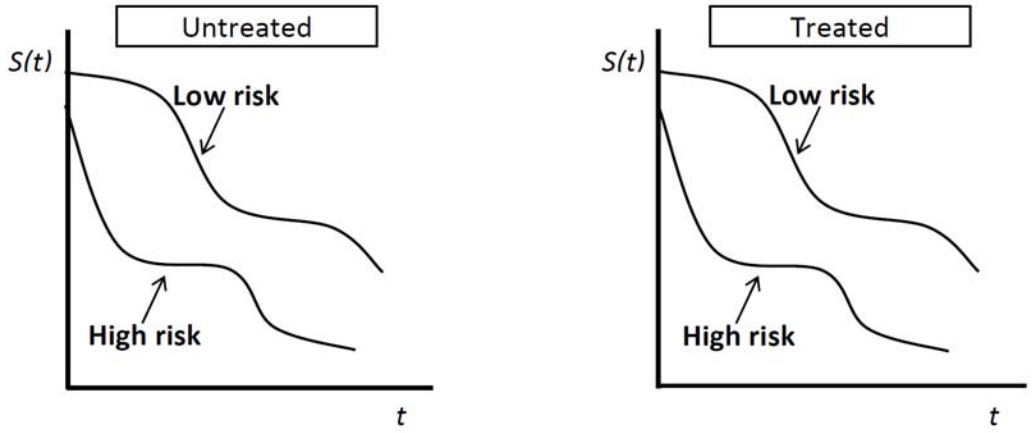
1.8 Predictive versus Prognostic Tools in Breast Cancer

As previously mentioned, breast cancer is a heterogeneous and phenotypically diverse disease. There are multiple biologic subtypes of breast cancer and they each have a distinct behaviour and response to therapy. Classical pathologic and clinical indicators have been identified as factors that predict the prognosis of a patient with breast cancer. Such factors include age, patient co-morbidity, tumour size, the presence of lymphovascular invasion, high grade and poor-differentiation of a tumour as well as nodal involvement.

However, recently gene expression arrays have been used to identify profiles associated not only with good and poor outcomes in breast cancer patients, but also with response to specific therapies, such as chemotherapy or anti-estrogen treatment. In our current therapeutic model, treatments are not tailored specifically to the individual. There are general guidelines for a specific kind of tumour, however, we are unable to predict before treatment if patients will benefit or not from the treatments they receive. This leads to some patients being over-treated and incurring toxicities needlessly while others are undertreated. Predictive tools based on genetic expression arrays, can therefore provide more successful tailored treatments for patients.

Predictive tools are based on the actual tumours' molecular make up to determine whether or not they would benefit from a particular therapy. On the contrary, general prognostic tools will classify patients based on their gene expression profile into groups of good or poor prognosis irrespective of whether they respond to treatment. Importantly, characterization of the prognostic or predictive value of a biomarker identified in patients undergoing a given treatment requires a control group of non-treated patients, or treated with an alternative drug acting via an unrelated mechanism. A pictoral representation of the distinction between a predictive versus a prognostic tool can be seen in Figure 3, p.30.

A.



B.

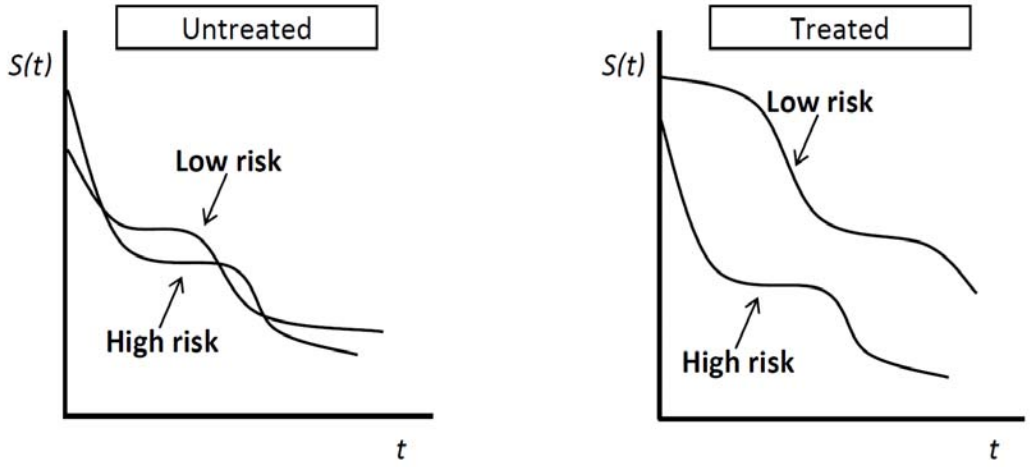


Figure 3: Prognostic versus Predictive tools

A= Prognostic tool (whether untreated or treated, patients are categorized into a high risk versus low risk group; **B** = Predictive tool (will show the likelihood of response to a specific treatment).

1.9 Predictive Tools in Breast Cancer

The two most common gene signatures used clinically today are the 21-gene recurrence score (Oncotype DX) [81] and the Amsterdam 70-gene prognostic profile (Mammaprint)[82]. Oncotype DX is a predictive tool whereas Mammaprint is a prognostic tool. Multiple studies have shown that although some patients with ER+ tumours derive a benefit from adjuvant chemotherapy, the majority do not. Oncotype DX was created to identify the subset or ER+, node negative patients who would derive benefit and thus avoid over-treating patients who did not. Based on the genetic profile of a patient's tumour, a recurrence score can be calculated and thus guide treatment. Mammaprint aids as a guide for decision-making with regards to adjuvant therapy. This tool also based on the genetic profile of a tumour, will categorize it as either being a good prognosis or poor prognosis tumour, once again guiding clinical decision-making.

1.10 Development of a gene signature that can predict the response to Tamoxifen

As already mentioned, two thirds of breast cancers are ER+ and their growth is stimulated by estrogens. Adjuvant therapy with anti-estrogens such as Tamoxifen and AIs has been shown to increase survival in breast cancer patients. This treatment is, however, not successful in all ER-positive tumours, with up to 40% of patients recuring despite completed treatment. Tumours can present intrinsic or acquired resistance to Tamoxifen, the mechanisms of which were described earlier. However, it is currently impossible to predict which patient will benefit from Tamoxifen therapy and which will not.

Preliminary studies in Dr. Mader's lab have identified 20 genes whose expression levels in tumours are able to predict the response to Tamoxifen therapy (disease-free survival

including local-regional recurrence and metastatic recurrence). These markers, identified using bioinformatics analysis of published gene expression datasets, were able to discriminate patients that would respond best to Tamoxifen from those that did not.

DNA microarray assays performed on the ER+ MCF-7 breast cancer cell line, allowed Dr. Mader's laboratory to identify 170 primary estrogen receptor target genes [83]. Bioinformatics tools were thereafter used to demonstrate, using several published datasets of breast tumour expression profiles, that levels of expression of these genes in patients' tumours predict outcome for Tamoxifen-treated breast cancer patients. A 20-gene signature was then derived from the best estrogen primary target genes combined with genes from unrelated signalling pathways found to have individual predictive value and was found to have a superior predictive value for Tamoxifen efficacy when tested against the 170-gene model[84],[85],[86]. This signature's predictive value was found to be robust in two tumour gene expression datasets (Desmedt C *et al.* 2007[87], Sortiriou *et al.* 2006[88]) and independent from traditional predictors such as ER/PR and lymph node status (Figure 4, p. 33). This figure demonstrates the heatmap and Kaplan-Meier plots for the 20-gene signature.

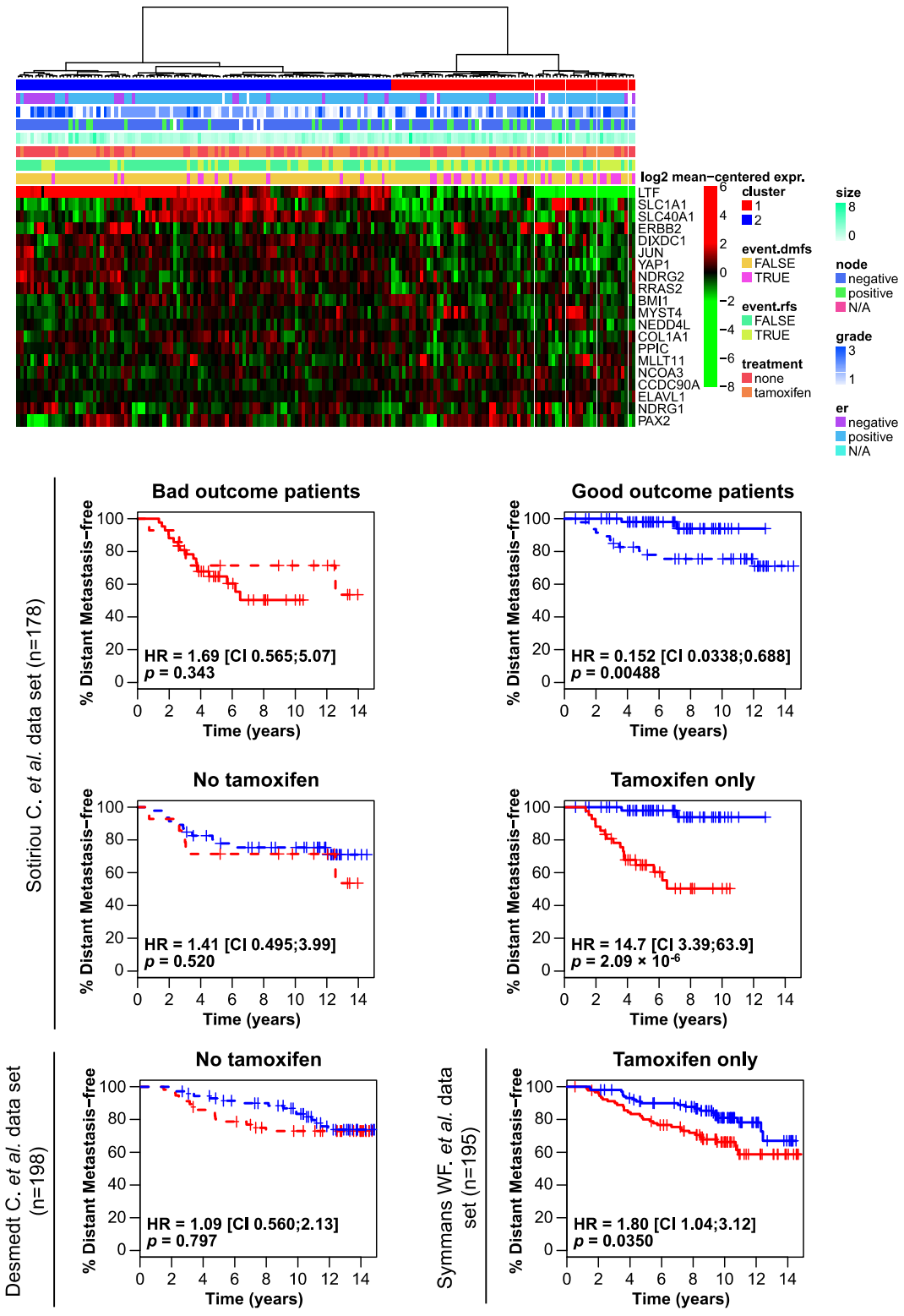


Figure 4: Heat map of the 20-gene signature applied to several published data sets demonstrating its ability to predict the response to Tamoxifen

The overall purpose of this study is to develop a PCR kit (gene signature) to monitor expression levels of these 20 genes and ultimately to test this 20-gene signature in a retrospective study using paraffin-embedded breast cancer tissues of patients with a known medical history. This tool will thus have a direct impact on clinical practice through the development of markers of therapeutic success for treatment with Tamoxifen and possibly AIs. Futile treatments would be avoided, preventing needless side effects, and improved identification of ER+ tumours with a low chance of success to anti-estrogen therapy will facilitate research into more appropriate treatments for hormone resistant tumours.

1.11 Identified Predictive Genes

The ultimate goal of our study was to develop a 20-gene signature as seen in the preliminary results. However, these genes were identified using data sets of Affymetrix micro-arrays. Knowing that Q-PCR assays may not always reproduce results obtained with micro-array probes, 10 additional robust predictive genes were identified for validation, so as to ensure a final 20-gene signature in the instance where genes could not be validated in Q-PCR.

Good predictors

There were 15 genes that were selected as predictors of a good prognosis and good response to Tamoxifen therapy. A list of these genes as well as a summary of their function can be found in Table II, p. 35.

Table II: Genes that predict a good response to Tamoxifen

Gene	Name	Function
ABAT	Aminobutyrate Aminotransferase	Catalyses conversion of GABA into succinic semialdehyde
COL1A1	Collagen type 1, alpha 1	Collagen found mainly in cartilage
LTF	Lactotransferrin	Inhibits solid tumor growth and metastasis
PPIC	Peptidyl-prolyl isomerase C	Catalyses a rate limiting step in protein folding
TNFRSF10B	Tumor Necrosis Factor Receptor Superfamily 10B	Apoptotic death of cancer cells
YAP1	Yes associated protein 1	Encodes protein that binds to SH3 domain of YES tyrosine kinase
RERG	Ras-like and Estrogen-regulated growth inhibitor	Inhibits cell proliferation and tumor formation
SLC1A1	Solute carrier family 1	Transports glutamate across plasma membranes
DIXDC1	Dix-domain containg protein 1	Inhibits axin-mediated JNK activity
EGR1	Early growth response 1	Inhibits human cancer cell growth
PAX2	Paired box gene 2	With ER coactivator NCOA3 compete for binding and regulation of ERBB2 transcription
SLC40A1	Solute carrier family 40, member 1	Duodenal iron export protein
BMI-1	Leukemia viral BMI-1 oncogene, mouse	Regulating proliferative activity of normal stem and progenitor cells
JUN	Jun proto-oncogene	DNA, protein, transcription factor binding
RERGL	Ras Superfamily	Cell proliferation

Some genes in this group are of particular interest. Many of them have interesting functions in breast cancer, potentially explaining why overexpression of these genes in an ER+ breast cancer may predict a better outcome with Tamoxifen treatment.

LTF has been shown to inhibit the growth of solid tumours and the development of experimental metastases in mice[89]. This was then further illustrated when evaluating primary breast tumours and their metastases, showing that LTF significantly decreased the metastatic potential in breast cancer by inhibiting and thus decreasing cellular motility[90].

TNFRSF10B is among the tumour-necrosis factor receptor superfamily, which is associated with its TNF-related apoptosis-inducing ligand (TRAIL) R2. This ligand induces the process of cancer cell death/apoptosis. Although one study showed that TRAIL R2 was associated with higher-grade tumours, when compared to TRAIL R1, both are involved in cancer cell apoptosis and mammary carcinoma could be sensitised to TRAIL-R2-induced apoptosis, suggesting that TRAIL-R2 might therefore be used to therapeutically target such tumours[91].

There is some suggestion in the literature that YAP1 is a tumour suppressor gene for breast cancer and thus if lost, may lead to more progression of breast cancer cells. In corollary therefore, if highly expressed, may lead to cancers with a better prognosis[92].

In a five-gene model predicting the outcome of patients with early ER+ breast cancer, RERG overexpression was associated with increased survival and better outcome in patients with ER+ cancers treated with Tamoxifen[93]. This positive correlation has been noted in other studies, associating RERG with a better prognosis in breast cancer patients. One study showed that high RERG expression correlated with the expression of a set of genes that defined the ER+ subtype and was associated with a slow rate of tumour cell proliferation and growth inhibition and thus a favourable prognosis for these cancer patients[94]. Moreover, RERG expression was inversely associated with the proliferation marker MIB1. Strong RERG

expression showed an association with longer breast cancer specific survival and distant metastasis free interval in a series of luminal type breast cancers and these associations were independent of other prognostic variables[95].

PAX2 has been associated with less invasive phenotypes and thus a better prognosis in breast cancer. It has been shown that PAX2 activation by estradiol is selectively achieved in breast cancer cells of the luminal subtype, via ERα, and identifies IGF-1 as a negative regulator of PAX2 activity in these cells. Further, a new role for PAX2 in the maintenance of a low invasive behavior in luminal breast cancer cells upon exposure to estradiol has been revealed, and shows that overexpression and activation of PAX2 in these cells is sufficient to reduce their invasive ability[96]. Another study showed that breast cancers overexpressing PAX2 were less likely to recur[97]. This may be due to PAX2 ability to repress ERBB2 and therefore lead to better outcome. PAX2 has been identified in a novel role, as a crucial mediator of ER repression of ErbB2 by the anti-cancer drug tamoxifen. PAX2 and the ER co-activator NCOA3/SRC-3 compete for binding and regulation of ErbB2 transcription, the outcome of which determines tamoxifen response in breast cancer cells. The repression of ErbB2 by ER-PAX2 links these two important breast cancer subtypes and suggests that aggressive ErbB2 positive tumours can originate from ER positive luminal tumours by circumventing this repressive mechanism. These data provide potential mechanistic insight into the molecular basis of endocrine resistance in breast cancer[98].

SLC40A1 has also been shown to predict a better outcome in breast cancer patients. SLC40A1 (ferroportin) exports iron out of the cell. Since malignant cells have a high demand for iron to grown, upregulation of SLC40A1 which exports iron out of the cell can lead to

growth inhibition and decreased risk of metastasis[99]. This relationship between SLC40A1 as an iron exporter and good prognosis, was reconfirmed in another study by Miller *et al.*[100].

The role of BMI-1 in breast cancer is uncertain with some studies relating it to a good prognosis and others with a poor prognosis. It is especially noted to be a marker of good prognosis in ER+ cancers[101]. However, another study noted that BMI expression was noted more often in grade 3 basal-like phenotype, which independently correlated with a worse prognosis[102]. It may be that BMI-1 may play a different role in different subtypes of breast cancer.

Finally, JUN (also known as AP1) has also been implicated in breast cancer to show a better prognosis. It has been shown that p12CDK2-AP1 over-expression inhibited in vivo tumor growth in immunodeficiency mice, supporting an inhibitory role for p12CDK2-AP1 in breast cancer development[103].

Poor predictors

There were 15 genes that were selected as predictors of a poor prognosis and poor response to Tamoxifen therapy. A list of these genes as well as a summary of their function can be found in Table III, p. 39.

Table III: Genes that predict a poor response to Tamoxifen

Gene	Name	Function
CCDC90A	Coiled Coil domain containing protein 90A	In mitochondria, protein of unknown function
ERBB2	Erythroblastic leukemia viral oncogene homolog 2	Cell surface receptor similar to EGFR, overexpression in Her-2 positive breast cancer
FOXM1	Forkhead box protein M1	Key role in cell cycle and proto-oncogene among many cancers including breast cancer
NCOA3	Nuclear Receptor Coactivator 3	Stimulates transcriptional activities in hormone dependent fashion, overexpressed in breast CA
PGK1	Phophoglycerate kinase 1	Secreted by tumour cells, causing inhibition of angiogenesis by release of angiostatin
PRIM1	Primase Polypeptide 1	A subunit of the DNA polymerase complex, initiation of DNA replication
RRAS2	Related Ras viral oncogene homolog 2	Oncogene
NEDD4L	Ubiquitin protein ligase NEDD4-like	Links ubiquitin dependent protein degradation to the replication-recombination machinery
TNC	Tenascin C	Lost as human breast cancer cells develop metastasic potential
ELAVL1	Embryonic lethal abnormal vision drosophilia homolog-like 1	Destabilise mRNA and play role in control of gene expression
NDRG1	N-MYC downstream regulated gene 1	Growth arrest and cell differentiation
NDRG2	N-MYC downstream regulated gene 2	Neurite outgrowth
MLLT11	Myeloid/Lymphoid translocated to 11	Function unknown
VDAC2	Voltage dependent anion channel 2	Regulates the activity of BAK and connection btw mitochondrial physiology and apoptosis
MYST4	Histone Acetyltransferase MYST4	Plays a role in positive and negative transcription regulation

Several of these genes have already characterized functions in breast cancer, potentially explaining why overexpression of these genes in an ER+ breast cancer may predict a poor outcome with Tamoxifen treatment.

The upregulation of ERBB2 (Her-2 overexpression) noted in ER+ breast cancers has been one explanation of acquired Tamoxifen resistance in ER+ breast cancers and may explain the poor outcome to Tamoxifen treated ER+ breast cancers that overexpress ERBB2[39],[104].

Elevated expression of FOXM1 in breast cancer has been shown to correlate with an undifferentiated tumour phenotype and thus a negative clinical outcome. One study showed its ability to act as a transcriptional repressor, playing an important role in regulating the differentiation of luminal epithelial progenitors. Regeneration of mammary glands with elevated levels of FoxM1 was shown to lead to aberrant ductal morphology and expansion of the luminal progenitor pool. Conversely, knockdown of FoxM1 resulted in a shift towards the differentiated state. FoxM1 mediates these effects by repressing the key regulator of luminal differentiation, GATA-3, by promoting methylation of the GATA-3 promoter in an Rb-dependent manner, identifying FoxM1 as a critical regulator of mammary differentiation with significant implications for the development of aggressive breast cancers[105]. In another study, silencing of FOXM1 abolished cell proliferation and overcame acquired Tamoxifen resistance suggesting that FOXM1 may contribute to anti-estrogen insensitivity[106].

Patients whose tumours show elevated expression of NCOA3 (also known as AIB1) have significantly shorter disease-free and overall survival times after surgery than other patients with breast tumours[107]. Another paper showed that AIB1 plays an important role in Tamoxifen resistance, and inhibiting AIB1 significantly restores the sensitivity of Tamoxifen

in ER-positive BT474 breast cancer cells[108]. Convincing clinical studies have shown that high levels of HER family member proteins have been associated with relapse after Tamoxifen therapy in breast cancer patients that have high AIB1 protein expression[109].

PGK1, a protein involved in glycolysis, was found to be upregulated in HER-2/*neu*-positive breast tumours. Increased glycolytic flow and energy production may contribute to the acquired resistance to Tamoxifen[110].

RRAS2 is a RAS family member that shares more than 50% amino acid sequence identity with classic RAS proteins (HRAS, NRAS, and KRAS). RRAS2 encompasses almost identical functional domains to these latter proteins and accordingly shares a number of effectors, including RAF1. A link between RRAS2 and Tamoxifen response has already been suggested. Silencing RRAS2 leads to sensitivity to Tamoxifen and specifically, high RRAS2 protein expression in breast tumour biopsies has been shown to correlate with a shorter time to relapse in Tamoxifen-treated patients, supporting the case that RRAS2 plays a critical role in determining the response to Tamoxifen.[111]. In patients treated with adjuvant Tamoxifen monotherapy, high cytoplasmic TC21 (also known as RRAS2) tumor expression has been shown to confer an increased recurrence rate. There is growing evidence that crosstalk between ER and growth factor signaling contributes to Tamoxifen resistance[112].

TNC is an adhesion-modulating extracellular matrix protein and is highly expressed in the microenvironment of most solid tumours including breast cancer and is frequently upregulated in a variety of pathological conditions including chronic inflammation and cancer. TNC has been implicated in the modulation of cell migration, proliferation, invasion and angiogenesis[113]. The protein has been shown to be involved in a wide variety of processes

such as proliferation, epithelial-to-mesenchymal transition, tumor cell migration, invasion, and metastasis. TNC is associated with Tamoxifen resistance and shorter metastasis free and progression free survival after adjuvant Tamoxifen treatment with lack of clinical benefit from first-line Tamoxifen monotherapy. High TNC expression could be a (indirect) marker for a defective estrogen response pathway, which would render the cells unresponsive to endocrine treatments such as Tamoxifen[114]. Moreover, breast cancer cells that infiltrate the lungs support their own metastasis-initiating ability by expressing tenascin C (TNC). It was found that the expression of TNC was associated with the aggressiveness of pulmonary metastasis. Cancer cell–derived TNC promotes the survival and outgrowth of pulmonary micrometastases[115]. Higher TNC expression was also shown in invading fronts of breast cancer, correlating with poorer patient outcome. It could induce EMT-like change showing loss of intercellular adhesion and enhanced migration in breast cancer cells, associated with FAK phosphorylation by SRC; this may be responsible for the observed promotion of TNC in breast cancer invasion[116].

Although some studies have shown that NDRG1 and NDRG2 may be associated with a good outcome, there are other studies that have shown the opposite. One study showed that NDRG1 overexpression was related to shorter disease free survival and a poor prognosis in luminal A and triple negative breast cancers[117].

MLLT11 (also known as AF1Q) overexpression enhanced the in vitro proliferation and invasive potential of breast cancer cells. In an *in vivo* study, it was demonstrated that AF1Q transfected breast cancer cells grew much faster and had more pulmonary metastases than vector-transfected or its parental counterparts. On the contrary, AF1Q knockdown cells grew slower and had less pulmonary metastasis. Taken together, these results provide functional

evidences that overexpression of AF1Q leads to more progression in human breast cancer, at least in part, through regulating the integrin a3, Ets-1, MMP-2, EFP, and 14-3-3d expression[118].

Finally, VDAC2 inhibits Bak1, which is activated by Bax in mitochondrial apoptosis. VDAC2 has also been associated with poor outcome in breast cancer especially in patients with supra-clavicular nodal metastases[119].

2. Hypothesis and Objectives

2.1 Hypothesis

The main hypothesis is that we can translate this microarray-based signature into a 20-gene signature PCR tool whose role in predicting the Therapeutic response to Tamoxifen will subsequently be tested using a retrospective cohort of patients.

2.2 Objectives

The main objective of this project is to test our hypothesis by developing PCR probes against each of these genes and validating their capacity to detect expression predicted by micrarray analysis using a panel of breast cancer cell lines. Expression in tumours will also be assessed to determine the robustness of the assay in this setting.

Specific aims:

(i) To optimise the method of RNA extraction from fresh tissue and formalin-fixed paraffin-embedded tumours (FFPE)

(ii) To correlate the expression levels of the predictive genes in breast cancer cell lines between QPCR and Microarray

(iii) To identify PCR probes that accurately monitor expression levels of our marker genes in eight different breast cancer cell lines, fresh tissue and FFPE tumours

(iv) To determine which housekeeping genes are the most reliable to include in our tool

(v) To create a low density array card with 20 predictive genes and 4 housekeeping genes

3. Materials and Methods

3.1 Gene selection

Gene selection was performed, with standard bioinformatics tools using published data sets of breast cancer expression profiles (Affymetrix microarrays) by Slim Fourati, a PhD candidate in Dr. Mader's lab. An overview of how this was done is provided below.

Initially, tumours were subjected to a training screen which compared Tamoxifen treated patients versus patients who did not receive any adjuvant systemic therapy, using 3 published data sets (Miller LD. *et al.*, 2005; Sotiriou C. *et al.*, 2006; Chin K. *et al.*, 2006)[88],[120], ,[121] . Cox proportional regressions were used to identify genes associated with distant metastasis-free survival of Tamoxifen-treated patients but not of untreated patients, and an interaction test was then performed to assess the statistical significance of this association. Genes with a p-value equal to or less than 5% were considered predictors of Tamoxifen response. Genes predictive in all 3 data sets were included in the validation process. There were a few exceptions to this rule as they were felt to be good candidate genes. TNC, RRAS2, PAX2, NCOA3 and LTF were not found in all the training screen datasets, however were shown to be predictive in the literature with QPCR in Tamoxifen-treated versus untreated patients and thus were kept as candidate genes that merited further study.

The validation process took the genes that passed the screening test and were then applied to 3 other published data sets of Tamoxifen treated patients (Loi S. *et al.*, 2007, Loi S. *et al.*, 2008, Zhang Y. *et al.* 2009)[122],[86],[123]. Cox proportional regressions were fit to the expression of all candidate genes (identified in the training step) and a likelihood ratio test was

used to assess the statistical significance of the association of the candidate genes and distant-metastasis free survival of tamoxifen treated patients. Genes that remained predictive in at least 2 out of 3 datasets were considered validated. From this process 30 predictive genes were identified. Table IV (p. 48) describes these genes and details results obtained during the selection process.

Table IV: Identified Predictive Genes

Gene	Name	Outcome Predictor	Training Screen
ABAT	Aminobutyrate Aminotransferase	Good	All 3 data sets
COL1A1	Collagen type 1, alpha 1	Good	All 3 data sets
LTF	Lactotransferrin	Good	1 data set
PAX2	Paired box gene 2	Good	2 data sets
PPIC	Peptidyl-prolyl isomerase C	Good	All 3 data sets
TNFRSF10B	Tumor Necrosis Factor Receptor Superfamile 10B	Good	All 3 data sets
YAP1	Yes associated protein 1	Good	All 3 data sets
RERG	Ras-like and Estrogen-regulated growth inhibitor	Good	All 3 data sets
SLC1A1	Solute carrier family 1	Good	All 3 data sets
DIXDC1	Dix-domain containing protein 1	Good	All 3 data sets
EGR1	Early growth response 1	Good	All 3 data sets
SLC40A1	Solute carrier family 40, member 1	Good	All 3 data sets
BMI-1	Leukemia viral BMI-1 oncogene, mouse	Good	All 3 data sets
RERGL	Ras Superfamily	Good	All 3 data sets
JUN (AP1)	Jun proto-oncogene	Good	All 3 data sets
CCDC90A	Coiled Coil domain containing protein	Poor	All 3 data sets
ERBB2	Erythroblastic leukemia viral oncogene homolog 2	Poor	All 3 data sets
FOXM1	Forkhead box protein M1	Poor	All 3 data sets
NCOA3	Nuclear Receptor Coactivator 3	Poor	2 data sets
PGK1	Phosphoglycerate kinase 1	Poor	All 3 data sets
PRIM1	Primase Polypeptide 1	Poor	All 3 data sets
RRAS2	Related Ras viral oncogene homolog 2	Poor	1 data set
NDRG1	N-MYC downstream regulated gene 1	Poor	All 3 data sets
NDRG2	N-MYC downstream regulated gene 2	Poor	All 3 data sets
NEDD4L	Ubiquitin protein ligase NEDD4-like	Poor	All 3 data sets
TNC	Tenascin C	Poor	No data sets
ELAVL1	Embryonic lethal abnormal vision drosophilia homolog-like 1	Poor	All 3 data sets
MLLT11	Myeloid/Lymphoid translocated to 11	Poor	All 3 data sets
VDAC2	Voltage dependent anion channel 2	Poor	All 3 data sets
MYST4	Histone Acetyltransferase MYST4	Poor	All 3 data sets

3.2 Cell Line Selection

Eight breast cancer cell lines were selected to evaluate the expression of the identified genes. The cell lines included ER positive lines such as MCF-7, BT-474, T47D, and ZR-75 as well as ER negative and Her-2 positive cell lines such as MCF-10F, MDA-MB-231, BT-20 and SKBR3. This allowed us to determine the variability of the expression of these genes in different breast cancer cell lines, to ensure that there is indeed a distinct difference between the molecular markers noted in an ER-positive versus an ER-negative or Her-2 positive cancer. Table V (p. 49) lists the media that cell lines were cultured in. Cell lines were not treated with hormones.

Table V: Cell lines and culture medium

Cell Line	Culture Medium
MCF-7	AMEM (with 20 nM of L-glutamine) 10% FBS
BT-474	DMEM 10% FBS
T47D	DMEM 10% FBS
ZR-75	RPMI 10% FBS
MCF-10F	DMEM/F-12 10% FBS, 10 ng/mL EGF, 10 µg/mL insulin, 0,5 µg/mL hydrocortisone, 100 ng/mL cholera toxin
MDA-MB-231	DMEM 5% FBS
BT-20	AMEM (with 20 nM of L-glutamine) 10% FBS
SKBR3	DMEM 10% FBS

3.3 RNA extraction

Cell lines

As previously mentioned, 8 cell lines were selected to evaluate the expression of the candidate genes in a variety of subtypes of breast cancer. RNA extraction was performed using Trizol Reagent®. The RNA extraction procedure from cell lines can be found in Appendix A.

Fresh Tissue

RNA extraction from fresh tissue (human breast cancer samples), which was preserved in RNA later, was also performed. Initially this was performed using an ABI protocol (Appendix B). However, the RNA yield was extremely low and impure. The main issue with this protocol was an inability to homogenize the tissue with mortar and pestle after having been placed on dry ice and frozen with liquid nitrogen as indicated by the protocol. Despite arduous manual grinding of the tissue to powder, the filter used in the purification step would get blocked, leading to loss of most of the RNA. A mechanical homogenizer may have facilitated this step and improved RNA extraction.

We therefore turned to the QIAzol Protocol, already used by the CHUM (Centre Hospitalier de l'Université de Montréal) Research Centre. Tissues were homogenized with a rotor-stator homogenizer (Dako Medimachine, Appendix C), leading to a much more successful RNA extraction and higher yield of material.

Formalin-fixed paraffin embedded tissue

RNA extraction from FFPE human breast cancer samples was also performed. Extracting RNA from FFPE can be challenging for many reasons. First, tissues that have been fixed for long periods of time have strong cross-linked bonds that are difficult to break, making isolation difficult. With archival FFPE tissues, time of fixation and the use of buffered versus non-buffered formalin as well as the age of the paraffin blocks are the main determinants of RNA quality[124]. Formalin fixation results in RNA degradation into small fragments and a low overall yield of RNA. Although microarray analysis or transcriptome sequencing is challenging, it is still possible to perform reliable Microarray as well as QPCR gene profiles with the RNA from FFPE samples[125], [126], [127]. A study by Leong *et al.* demonstrated that although RNA isolated from FFPE was relatively more degraded, 80% were still deemed suitable for subsequent assay (both DNA Microarray and RT-QPCR) and gene profiles generated were comparable to RNA extracted from paired fresh tissues[128].

Initially we used the ABI protocol for FFPE extraction (Appendix D). Yield was low as was quality of the RNA. An RNA clean up was thus performed however this did not improve the quality nor the yield of the RNA. Multiple adaptations were thus attempted to improve the quality and the yield of the RNA. It has been shown that longer digestion and more agitation may provide a higher yield of RNA from FFPE tissues[124]. I thus adapted the protocol by performing 20-25 minutes of rotation of the samples with xylene before placing it at 50°C for three minutes in the Deparaffinization step, as well as increased the digestion time to 3 hours at 50°C followed by an incubation at 80°C for 15 minutes. Other additional changes were made based on the ABI troubleshooting guide provided. In the Protease Digestion process, only 200 µL of digestion buffer was used as opposed to 400 µL. In the Nucleic Acid

Isolation process, only 240 µL (instead of 480 µL) of Isolation additive was added to each sample and 500 µL (instead of 1.1 mL) of 100% ethanol was added to the mix. Moreover, Wash 1 and 2/3 were centrifuged for 2 minutes as opposed to 30 seconds. Finally in the Nuclease Digestion process, since RNA yield was low, elution was done with 30 µL of RNase free water instead of 60 µL. This improved the 260/280 ratio but not the 260/230 ratio and the yield of RNA obtained remained low.

Due to this, I decided to compare the yield and quality of RNA extraction from FFPE tissues using two different protocols and kits (the same paraffin blocks were used in each protocol to directly compare them). The ABI protocol as previously mentioned was thus compared to the Roche Protocol of RNA Extraction from FFPE (Appendix E). Several specific changes were made to the protocols. Firstly, for both protocols, it was felt that the total quantity of the slices of the paraffin blocks (80 µm) as well as their thickness (4 x 20 µm slices) might have initially been too high to properly melt the wax and obtain good digestion. Thus instead of having 4 x 20 µm slices to equal 80 µm total, I decided to use less and thinner sections (10 x 5 µm slices equalling a total of 50 µm in each tube).

The main adaptation differences between the Roche Protocol and the ABI protocol can be found in Table VI (p.53).

Table VI: Differences between Roche and ABI Protocol for RNA extraction from FFPE tissues

Process	Roche	ABI
Deparaffinization	800 µL Hemo-De	1 mL Xylene
Digestion	Overnight incubation at 55°C	3 hour incubation at 55°C followed by 15 mins at 80°C
Drying after deparaffinization	Air dried at 55°C for 13 mins	Air dried at room temp for 45 mins

When comparing both protocols, the Roche Protocol had overall a higher yield of RNA as well as better quality RNA (especially with regards to the 260/230 ratio on the NanoDrop, see results section). It was also further noted that even with the Roche Protocol, higher yields still could be achieved if smaller quantities of starting material was used per tube. I therefore used 2 x 5µm slices per tube x 3 tubes for each sample (total tissue amount equalling 30 µm). The three tubes were then pooled together to increase the total yield of RNA obtained. Therefore the Roche Protocol provided better quality and higher yields of RNA, with less starting material from the paraffin block. Due to these findings, all RNA extractions from FFPE tissues were subsequently performed using the Roche Protocol for FFPE RNA extraction. The longer overnight incubation period, although more time consuming, may be instrumental in providing better RNA for expression profiling assays.

3.4 Quality Assessment of RNA

Once the RNA was extracted it was evaluated via two methods. First, using the NanoDrop® 2000 micro-volume spectrophotometer. Secondly, using the Agilent 2100 Bioanalyzer. These tools evaluate two quality control measures on isolated RNA. One is to

determine the quantity of RNA that has been isolated, the second is the purity and integrity of the isolated RNA.

NanoDrop®

The NanoDrop® was used to evaluate RNAs from all forms of tissue (cell lines, fresh tissue and FFPE tissue). It is easy to use and requires very little material. 1 μL of RNA is pipetted onto the pedestal and then the arm is closed. In a few seconds the RNA is analyzed and all three measures are calculated. Nucleic acids are traditionally quantified using UV absorption using a spectrophotometer. In its simplest form the absorbance is measured at 260 and 280 nm. The concentration of nucleic acid can be determined using the Beer-Lambert law, which predicts a linear change in absorbance with concentration. The OD at 260 nm is used to determine the RNA concentration in a solution, a A260 reading of 1.0 being equivalent to about 40 μg/mL of RNA. RNA has its absorption maximum at 260 nm and the ratio of the absorbance at 260 and 280 nm is used to assess the RNA purity of an RNA preparation. Pure RNA has an A260/A280 ratio of 1.8-2.1. Ideally a 260/230 ratio should be around 2.0[129].

Agilent BioAnalyzer

When using the Agilent Bioanalyzer, the following guidelines are used for the assessment of good quality RNA:

- sharp 18S and 28S peaks with no fragmentation peaks present
- a 28S/18S ratio of 2.0
- The sum of the total area under both peaks is at least 20% of the entire electropherogram
- Neither peak is greater than twice the size of the other
- RNA Integrity number (RNI) of 8-10 is considered pure RNA with good integrity

The BioAnalyzer was used to evaluate RNA extracted from cell lines. Since the RNA from FFPE tissue is by definition degraded, it would be inappropriate to test the quality of the RNA extracted from FFPE tissues using the BioAnalyzer.

3.5 Reverse Transcription

Once RNA is extracted it must be converted to cDNA for use in a Q-PCR. This is done via reverse transcription, which can be primed with either an OligoDT or a Random Hexamer. Because mRNA has a polyA tail, OligoDT will attach to the polyA tail and prime a full-length copy of the mRNA. For longer messages, a random primer is preferred as it will enable reverse transcription of the 5′ ends of long genes, but the resulting cDNAs may not be full-length copies of the entire gene.

Cell Lines

The exact protocol of reverse transcription that was used for the cell lines can be found in (Appendix F). For the cell lines Random Hexamer was used for reverse transcription.

Fresh Tissue and FFPE tissue

For the fresh and FFPE tissue, Gene Specific primers were used using an ABI protocol (Appendix G). Gene Specific primers enhance sensitivity by directing all of the RT activity to a specific message instead of transcribing everything in the mix. This is especially helpful when small fragments of RNA are being used, which makes this most appropriate for RNA extracted from fresh tissue and FFPE tissue.

3.6 Q-PCR Probe Selection

ABI TaqMan assay probes have all been validated by ABI and contain a 3' oligonucleotide, a 5' oligonucleotide and a fluorescent marker. They are the only assays that can be used on the Low-density array that we have chosen as the platform for our predictive tool. Different criteria were used to select the appropriate probes that would best represent our 30 genes. First, since cDNAs are more retrotranscribed at the 3'ends, Q-PCR TaqMan probes were chosen that were closest to the 3' end as possible. Second, because the RNA from patients is often composed of small fragments that are degraded when extracted from paraffin embedded samples, probes with a small amplicon length were preferred, ideally not exceeding 100-120 nt. Finally, gene selection was performed via bioinformatics tools using published data sets of breast cancer expression profiles (Affymetrix microarrays); a tool called the UMapit Microarray-to-TaqMan® Assays Mapping Tool by ABI was utilized allowing us to cross-map the TaqMan assays for our 30 genes that corresponded most closely to those used in the Affymetrix Microarray. Using these criteria, the ABI TaqMan assays were chosen and used to perform the Q-PCR experiments (Table VII, p. 57). It is important to note that there are no introns in the JUN gene, thus it was impossible to choose primers spanning an intron. It is noted that the probe used for MLLT11 had a longer amplicon of 142nt, however, it was the only probe available. This is most likely due to the fact the probe sequence of MLLT11 is unique (no matches with other regions of the genome) and it contains a "normal" frequency of GC nucleotides. Moreover, MLLT11 is a short gene (2.2 kb) thus less possibility of probes.

Table VII: TaqMan Assays used for the Q-PCR experiments

Target Acession Number	Gene Symbol	Closest Inventoried TaqMan Assay to 3' end	Amplicon length
NM_005180	BMI-1	Hs00180411_m1	105
NM_001031713	CCDC90A	Hs00978327_g1	94
NM_000088	COL1A1	Hs01076751_g1	59
NM_033425	DIXDC1	Hs00736707_m1	85
NM_001419	ELAVL1	Hs00171309_m1	75
NM_001005862	ERBB2	Hs99999005_mH	88
NM_002228	JUN	Hs01103582_s1	91
NM_002343	LTF	Hs00914330_m1	92
NM_006818	MLLT11	Hs00199111_m1	142
NM_012330	MYST4	Hs01043690_m1	71
NM_006534	NCOA3	Hs00180722_m1	59
NM_006096	NDRG1	Hs00608389_m1	69
NM_016250	NDRG2	Hs01045109_gH	90
NM_015277	NEDD4L	Hs00969321_m1	90
NM_000278	PAX2	Hs00240858_m1	57
NM_000943	PPIC	Hs00181460_m1	66
NM_012250	RRAS2	Hs00273367_m1	108
NM_004170	SLC1A1	Hs00188172_m1	76
NM_014585	SLC40A1	Hs00205888_m1	78
NM_006106	YAP1	Hs00902712_g1	62
NM_002160	TNC	Hs01115664_m1	87
NM_003375	VDAC2	Hs00748551_s1	94
NM_000291	PGK1	Hs00943178_g1	73
NM_202002	FOXM1	Hs00153543_m1	72
NM_000946	PRIM1	Hs01096422_g1	74
NM_032918	RERG	Hs00262869_m1	85
NM_024730	RERGL	Hs00226861_m1	72
NM_000663	ABAT	Hs00609436_m1	69
NM_003842	TNFRSF10B	Hs00366278_m1	62
NM_001964	EGR1	Hs00152928_m1	72

3.7 Q-PCR

The Protocol used for every Q-PCR experiment was an ABI Protocol (Appendix H). The experiments were done using a 24 x 16 well plate (384 well plate). Analysis was performed using the SDS 2.2.2 software of the 7900HT Q-PCR machine in the genomics platform of the IRIC. The results were then analysed and interpreted with the help of the SDS 2.2.2 software guidelines (Appendix I). All QPCR curves were individually analyzed. Samples were studied in triplicate format. Poor curves were eliminated from the analysis.

3.8 Selection of Housekeeping genes

Housekeeping genes are typically constitutively expressed genes that are required for the maintenance of basic cellular function, and are expressed in all cells of an organism under normal and pathophysiological conditions. These are included in a multi-gene PCR assay as they allow normalisation of the mRNA levels between different samples. However, the expression level of these genes may vary among tissues or cells and may change under certain circumstances.

A Q-PCR was performed using the ABI TaqMan Assays of 7 different housekeeping genes (PPIA, YWAHZ, ACTB, GAPDH, TBP, 18S and HPRT) tested on the eight different breast cancer cell lines. A software program called geNorm analyzed the results. After rigorous evaluation (as will be discussed in the results sections), 4 housekeeping genes were selected for our tool. A minimum of 3 housekeeping genes is required for reliable results. GeNorm is a popular algorithm to determine the most stable reference (housekeeping) genes from a set of tested candidate reference genes in a given sample panel. From this, a gene

expression normalization factor can be calculated for each sample based on the geometric mean of a user-defined number of reference genes[130].

3.9 Creation of a customized Low Density Array

The following is an overview of the experiments that led to the creation of the Low Density Array (LDA) card. Details and findings about these experiments will be further discussed in the Results section. The first step was to perform a QPCR using 7 housekeeping genes in the 8 selected breast cancer cell lines. This allowed us to select the 4 best housekeeping genes based on the lowest varia. The next step was to perform a QPCR using the 30 predictive gene TaqMan Assays on the 8 selected breast cancer cell lines. These results were then correlated to the Affymetrix Microarray, leading to the elimination of 5 genes. Using cDNA (from Gene specific RT), the remaining 25 genes were then tested using the TaqMan Assays on RNA extracted from paired FFPE and Fresh tissues of the same tumour. These genetic profile expressions were then correlated. These results led to the elimination of 5 more genes based on variability. This left 20 gene predictors and 4 housekeeping genes to include on the final LDA card, our ultimate predictive tool.

We therefore decided on the 48-gene card (24 genes in duplicates), which can be pre-loaded with the 20 TaqMan assays of our predictive genes and the 4 TaqMan assays of our housekeeping genes. As per ABI's recommendations, all cards have the 18S housekeeping gene occupying one of the ports as part of the tool. Please see Appendix J for the Q-PCR Protocol of an LDA card.

4. Results

4.1 Selection of Housekeeping Genes

RNA extraction was performed on 8 different breast cancer cell lines in duplicates (N1 and N2 replicates). Table VIII (p.60) describes the quantity and quality of the RNA extractions according to the NanoDrop and Agilent BioAnalyzer. Quantity, quality and integrity of RNA were all satisfactory for subsequent QPCR analysis.

Table VIII: Quantity and Quality of RNA extraction from Cell Lines (N1, N2)

Cell Line	N	Quantity RNA (ng/µl)	A260/280	A260/230	RIN
MCF-7	N1	411.81	1.90	1.89	9.5
	N2	454.20	1.91	2.03	8.8
BT-474	N1	408.92	1.89	2.15	8.3
	N2	370.48	1.88	2.31	8.4
T47D	N1	426.80	1.91	1.97	8.2
	N2	426.87	1.89	2.09	9.1
ZR-75	N1	415.98	1.88	2.07	9.4
	N2	406.48	1.86	2.20	9.3
MCF-10F	N1	346.84	1.92	1.36	9.2
	N2	455.15	1.88	2.09	9.3
MDA-MB-231	N1	364.01	1.91	2.01	9.7
	N2	393.81	1.86	2.31	9.4
BT-20	N1	397.80	1.93	1.64	7.7
	N2	304.81	1.86	2.12	10
SKBR3	N1	363.61	1.94	1.95	9.5
	N2	434.16	1.86	2.27	9.4

NanoDrop Results = Quantity RNA, A260/280, A260/230; RIN = RNA Integrity Number determined by Agilent BioAnalyzer

The purpose of the first QPCR experiment was to select the 4 most appropriate housekeeping genes to be used for all subsequent QPCR experiments. A Q-PCR was performed using the ABI TaqMan Assays of 7 different housekeeping genes (PPIA, YWAHZ, ACTB, GAPDH, TBP, 18S and HPRT) tested on the eight different breast cancer cell lines. The GeNorm program was then used to analyze the QPCR results. This experiment was performed twice (N1,N2) to assess technical variability. The housekeeping genes with a gene expression normalization factor, as calculated by the GeNorma alogotrihm, closest to 1 are the most stable (see Materials and Methods). Figures 5 and 6 (p. 62-63) illustrate the value each housekeeping gene received as well as a diagram describing which housekeeping genes were most stable for both the N1 and N2 experiments.

geNorm PrimerDesign

1.5	PPIA	YWAHZ	ACTB	GAPDH	TBP	18S	HPRT	Normalisation Factor
MCF10F	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.1223
MM231	1.56E+00	4.09E-01	8.81E-01	2.23E+00	1.80E+00	1.03E+00	9.71E-01	1.2603
SKBR3	8.34E-01	8.74E-02	3.52E-01	2.06E+00	1.71E-01	1.13E+00	4.87E-02	0.3787
BT20	1.93E+00	1.13E+00	7.69E-01	2.11E+00	1.14E+00	1.08E+00	1.65E+00	1.4865
T47D	1.40E+00	1.10E+00	8.48E-01	1.07E+00	7.85E-01	1.06E+00	9.63E-01	1.1410
ZR75	7.56E-01	1.88E-01	3.09E-01	3.80E+00	3.92E-01	1.12E+00	1.39E-01	0.5829
MCF7	1.93E+00	2.46E+00	6.93E-01	2.12E+00	1.45E+00	1.09E+00	1.61E+00	1.6918
BT474	2.37E+00	1.34E+00	5.57E-01	1.67E+00	8.83E-01	9.75E-01	3.75E-01	1.1161
M < 1.5	**0.954**	**1.343**	**0.952**	**1.365**	**0.990**	**1.098**	**1.444**	

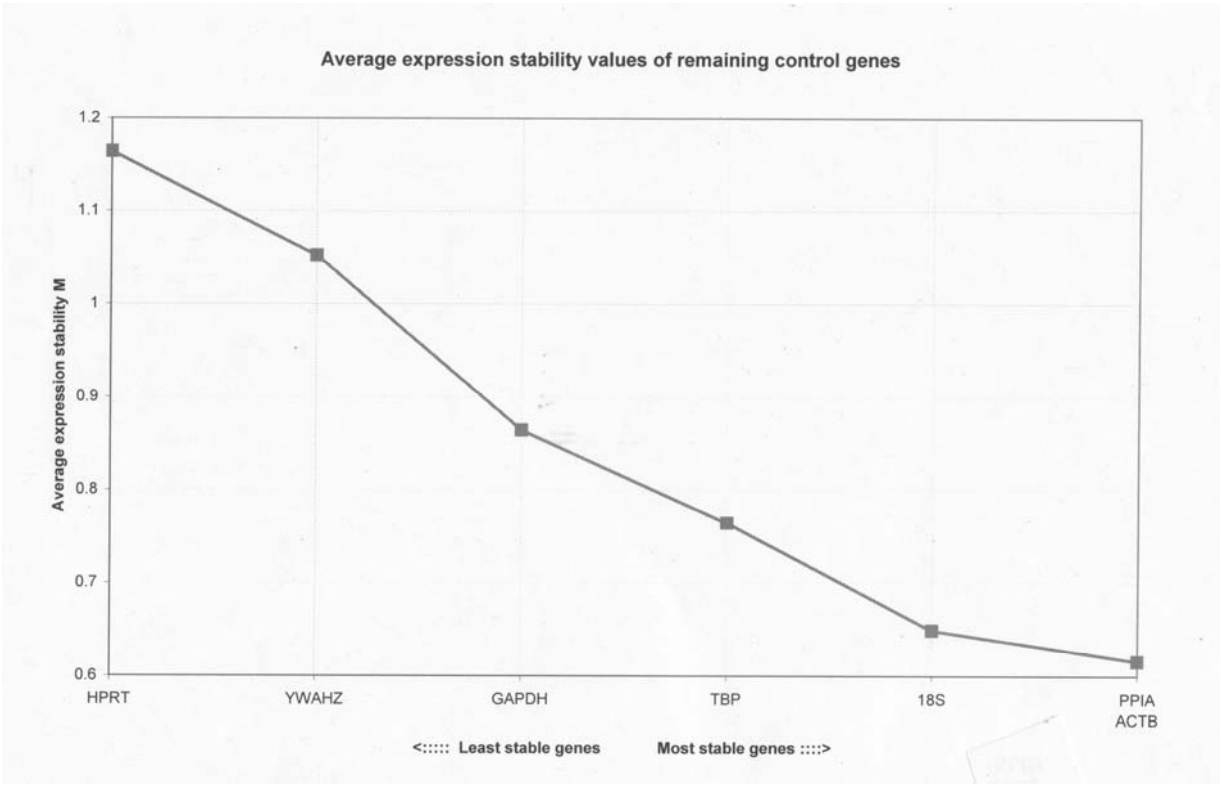


Figure 5: GeNorm Results (N1)

geNorm PrimerDesign

1.5	PPIA	YWAHZ	ACTB	GAPDH	TBP	18S	HPRT	Normalisation Factor
MCF10F	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.0830
MM231	1.22E+00	1.03E+00	1.10E+00	1.50E+00	2.00E+00	9.41E-01	1.28E+00	1.3638
SKBR3	1.08E+00	5.95E-01	5.16E-01	2.02E+00	4.61E-01	9.52E-01	1.26E-01	0.6764
BT20	6.31E-01	6.00E-01	9.05E-01	1.98E+00	4.52E-01	1.02E+00	5.42E-01	0.8410
T47D	9.27E-01	1.86E+00	1.07E+00	9.04E-01	8.48E-01	8.68E-01	5.43E-01	1.0222
ZR75	5.26E-01	4.27E-01	4.38E-01	2.93E+00	4.13E-01	1.03E+00	1.65E-01	0.6200
MCF7	7.65E-01	4.04E+00	5.15E-01	1.32E+00	1.19E+00	9.02E-01	1.25E+00	1.2554
BT474	1.99E+00	2.97E+00	9.00E-01	1.96E+00	1.43E+00	9.64E-01	6.72E-01	1.4961
M < 1.5	0.834	1.137	0.846	1.148	0.845	0.855	1.169	

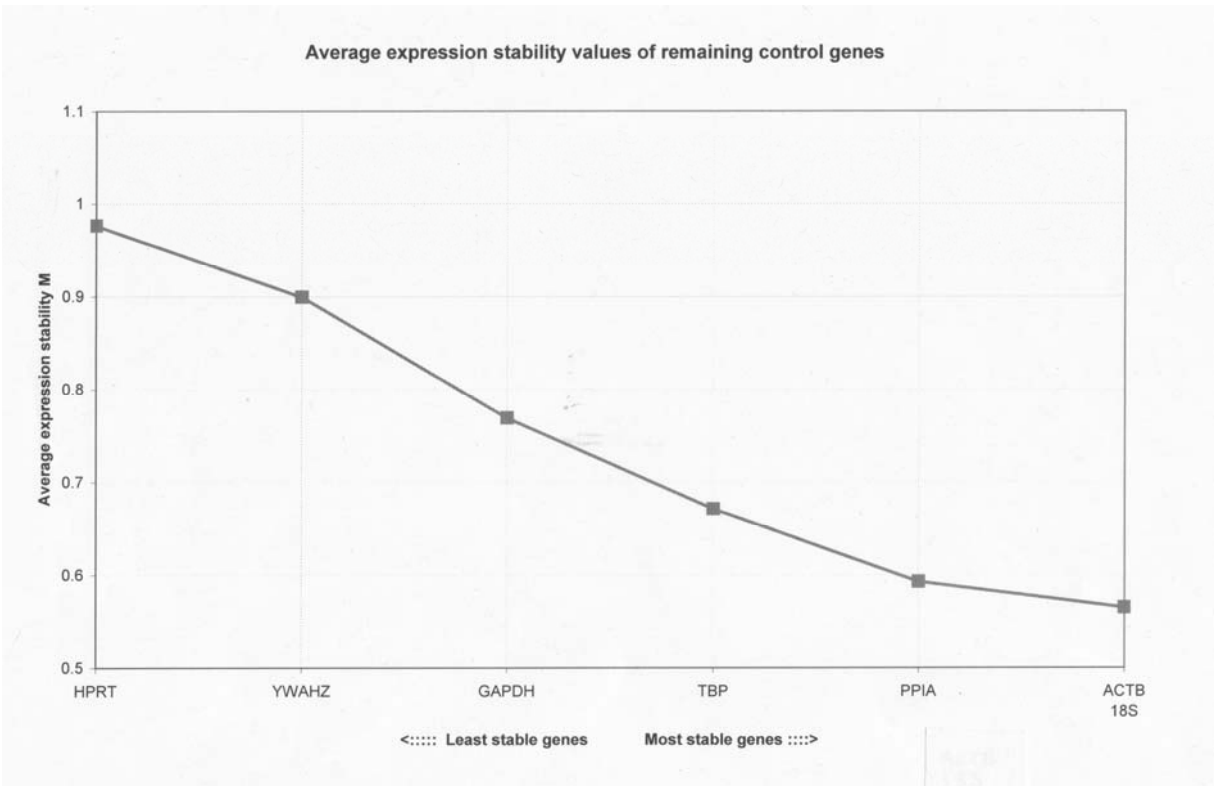


Figure 6: GeNorm Results (N2)

The most stable housekeeping genes were ACTB, 18S, PPIA and TBP. Table IX (p. 64) describes these genes and their function. These 4 genes were thus the best candidates to be used as the housekeeping genes in the subsequent QPCR experiments and eventual LDA. In order to have reliable results a minimum of 3 housekeeping genes must be used per experiment.

Table IX: Housekeeping Genes

Gene	Name	Function
18S	Ribosomal RNA	Ribosomal RNA subunit
ACTB	Actin Beta	Mammalian cytoplasmic non muscle actin, involved in cellular motility, structure and integrity
PPIA	Peptidyl-prolyl isomerase A	Protein folding, intracellular protein transport
TBP	Tata-box binding protein	Transcription factor

4.2 Selection of Predictive Genes

Using the ABI TaqMan Assays, the 30 predictive genes were evaluated by Q-PCR in the 8 different cell lines. Each sample was evaluated in a triplicate fashion. All Q-PCR curves were evaluated and poor curves were eliminated from the final analysis. One cell line was randomly selected as the calibrator sample to which other samples were compared. In this case, MDA-M231 was used as the calibrator sample. The relative quatification (RQ) is the fold change. The calibrator is fixed at a value of 1. The other samples have a value that is either greater or lower with regards to the calibrator. Almost all genes had a fold change of greater than 2 or less than 0.5, which is considered significant (Appendix I). The following Heat Maps describe the variations in expression of the 30 predictive genes in the 8 different cell lines (Figure 7 (N1) and 8 (N2), p. 65-66). A log scale was used to better discriminate the variations in gene expression. Variability was observed between replicates, leading to different clustering of the cell lines by the expression levels of the 30 genes. Note that the N2 replicate heatmap divides cell lines by ER status (with the exception of BT-20) and thus appears to represent more accurately the biology of the tumors.

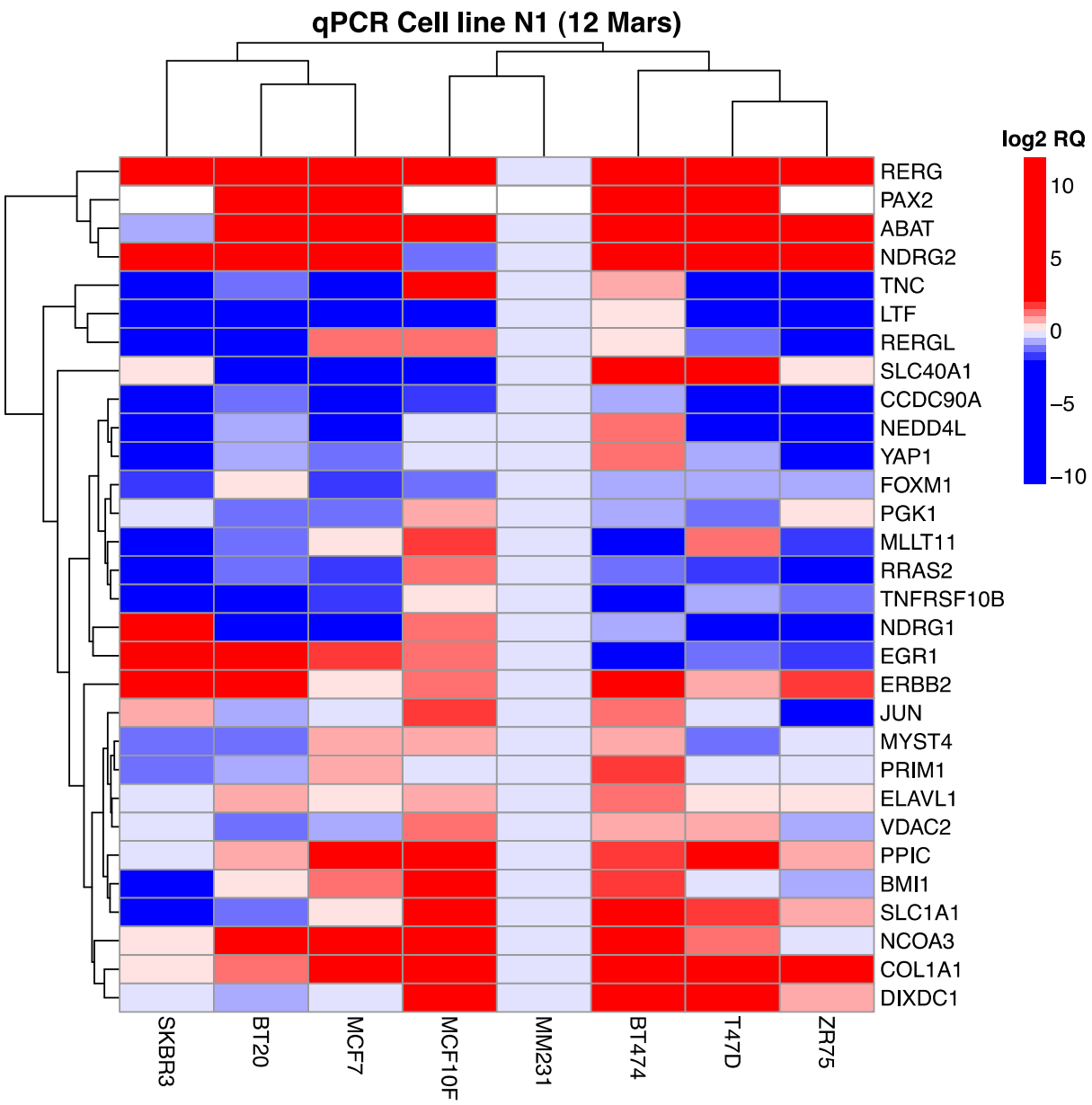


Figure 7: Heatmap results of the 30 Predictive genes in the 8 cell lines (N1)

Legend : Heat representation of the expression of 30 genes in breast cancer cell lines. The level of expression of each gene in each sample is represented in a log-space using a pink-blue colour scale. Rows of expression matrix, transcripts; columns, profile cell line. Hierarchial clustering (distance : correlation, linkage : ward) was used to regroup genes with similar expression patterns

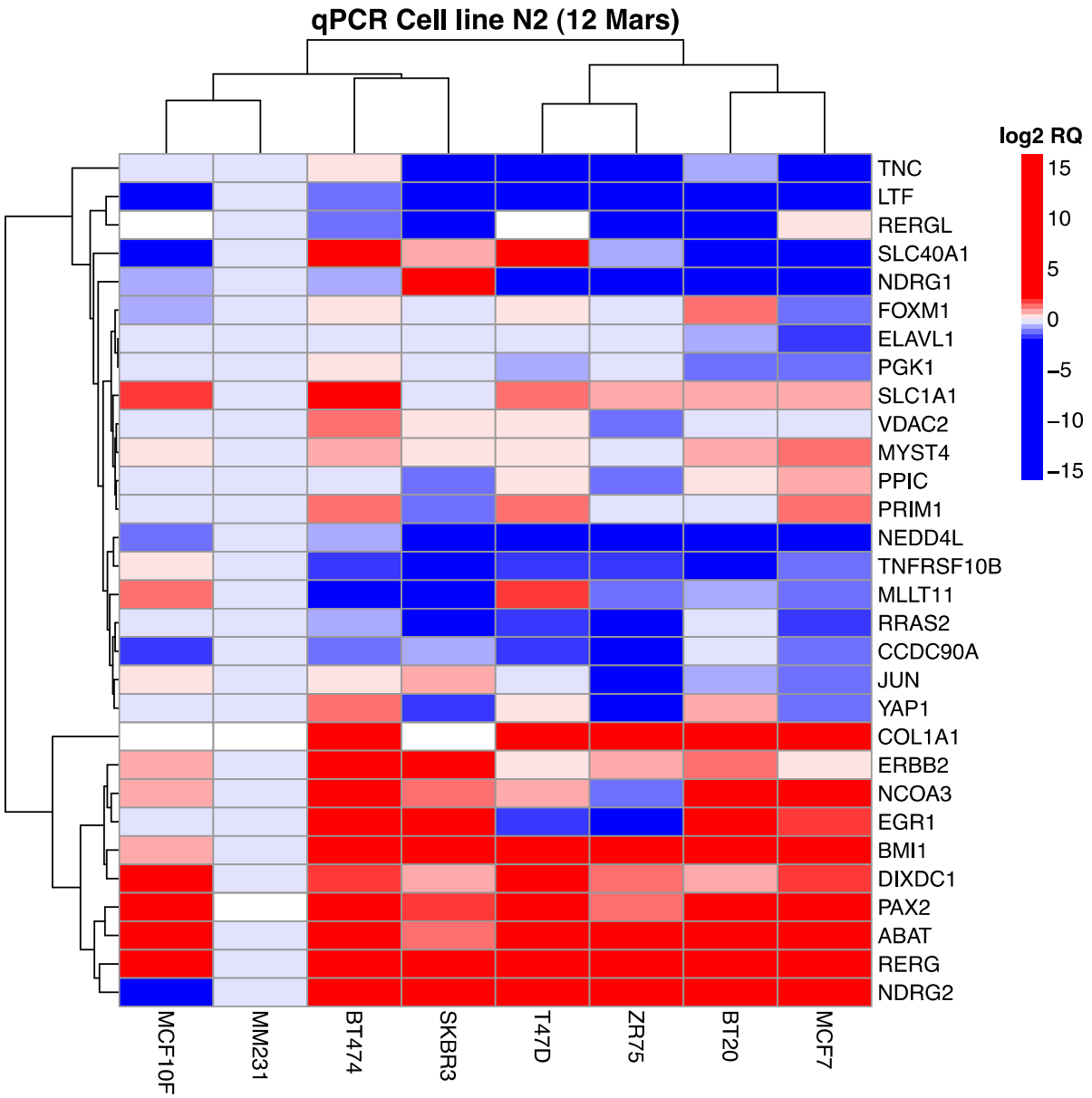


Figure 8: Heatmap results of the 30 Predictive genes in the 8 cell lines (N2)

Legend : Heat representation of the expression of 30 genes in breast cancer cell lines. The level of expression of each gene in each sample is represented in a log-space using a pink-blue colour scale. Rows of expression matrix, transcripts; columns, profile cell line. Hierarchial clustering (distance : correlation, linkage : ward) was used to regroup genes with similar expression patterns.

We then compared the patterns of expression obtained using ABI TaqMan Assays (N1 and N2 profiles) with 5 published Affymetrix microarray expression profiles in breast cancer

cell lines including the same 8 cell lines used in Q-PCR (HoeflichKP_2009, HollestelleA_2009, NeveRM_2006, LiedtkeC_2009, CharafeJauffretE_206); a heat map representation of the Spearman correlation between PCR and microarray for each gene is represented in Figure 9 (p. 68). We observe for most of the genes a positive correlation (red colour) between the QPCR experiments (N1 and N2 replicates) and the microarrays. 4 genes were eliminated due to a poor correlation with the Affymetrix microarrays: BMI-1, VDAC2, RERGL and MYST4. JUN was also eliminated on the basis of the absence of introns in this gene, leading to possible contamination by unspliced RNA precursors or by genomic DNA. Although PAX2 did not display a good correlation, it remained an interesting gene to study further due to its correlation with good outcome in the literature, and thus we decided to keep evluating this gene in subsequent experiments. Thus overall, 5 genes were eliminated from subsequent QPCR studies, based on this correlation analysis.

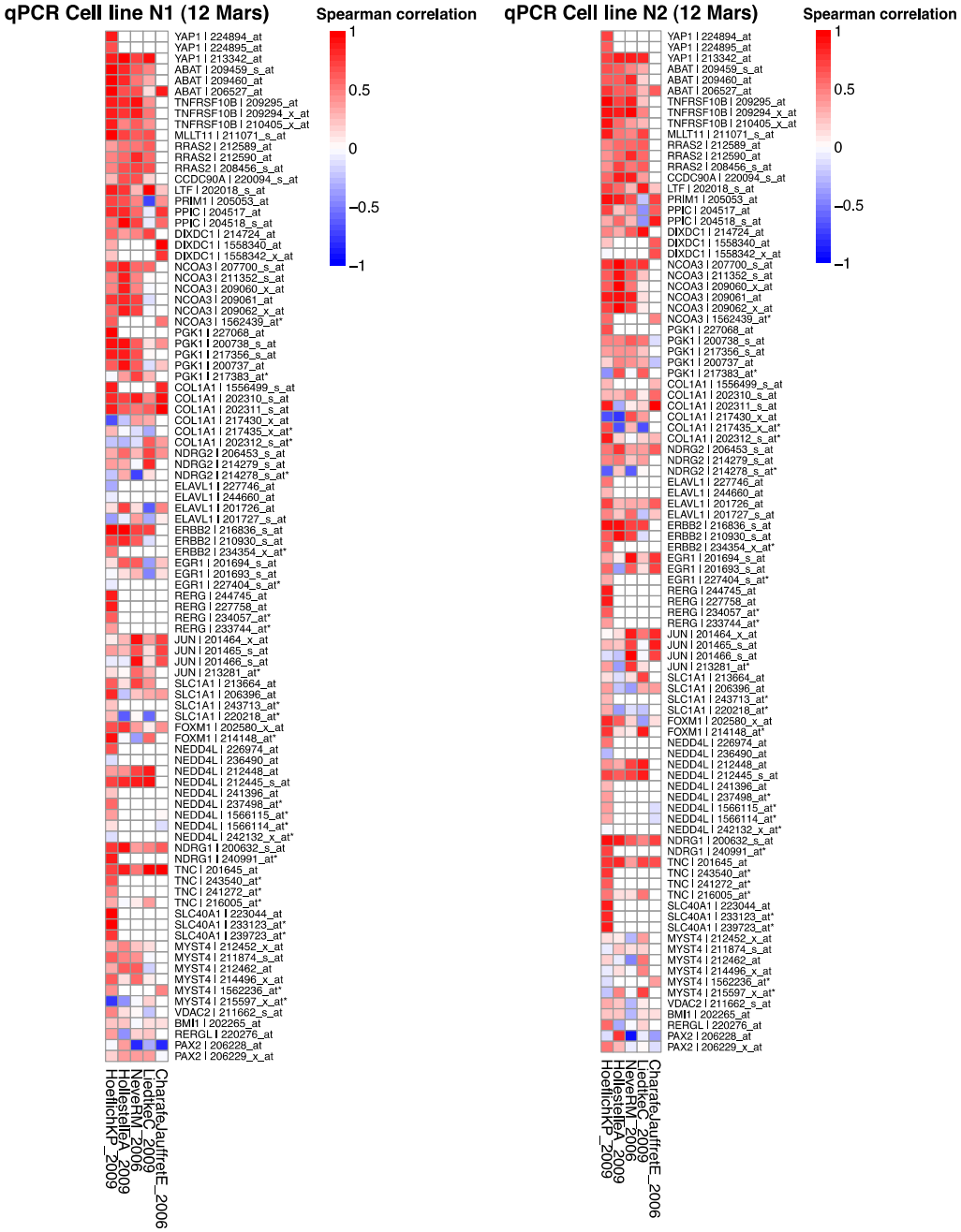


Figure 9: Correlation between ABI QPCR and Affymetrix Micro-Array of the 30 predictive genes on the 8 cell lines (N1 + N2)

Legend : Heatmap representation of the correlation of the PCR expression and the microarray expression across five breast cancer cell lines data sets. The Spearman corelation of each gene between PCR and microarray is represented in a log-space using a pink-blue color scale. Rows of expression matrix, transcripts; columns, microarray data set. Affymetrix probesets matching partially the genes are noted with a star (*).

4.2 Assessment of expression levels in fresh tissue and FFPE samples

As discussed in the methods, extracting RNA from fresh tissue using the ABI protocol yielded poor results, especially because homogenization of tissue in this protocol was inadequate. RNA was then extracted using the QIAzol CHUM protocol where homogenization was performed with the Dako Medimachine which was much more successful. Table X (p.69) compares the extraction results of both protocols.

Table X: Quantity and Quality of RNA extraction from Fresh Tissue (ABI vs. CHUM QIAzol Protocol)

Fresh Tissue	Protocol	Quantity RNA (ng/µl)	A260/280	A260/230
Sample 1	ABI	113.40	1.30	0.93
	CHUM	214.94	2.00	2.01
Sample 2	ABI	3.13	2.43	0.1
	CHUM	332.14	1.90	1.94
Sample 3	ABI	14.98	3.95	0.15
	CHUM	653.65	1.98	2.01

RNA extraction from Formalin-fixed Paraffin Embedded (FFPE) tissues was then perfected. As mentioned in the methods, 2 different protocols (ABI and Roche) were used and compared as well as adapted to have the best RNA extraction possible. We used the Nanodrop to evaluate the quantity and quality of the RNA extracted. The A260/280 value is the most relevant parameter when evaluating the quality of RNA from FFPE tissues. First, we evaluated the RNA extraction using the ABI protocol. The yield and quality of the extraction were low and so an RNA clean up was performed, and the RNA re-evaluated. This however, did not improve the quantity or quality of the RNA. These results can be seen in Table XI (p.70).

Table XI: Quantity and Quality of RNA extraction from FFPE (ABI Protocol before and after RNA clean-up)

FFPE Tissue	ABI Protocol	Quantity RNA (ng/µl)	A260/280	A260/230
Sample 1	Pre clean-up	155.87	2.02	1.38
	Post clean-up	81.43	1.91	2.03
Sample 2	Pre clean-up	93.63	2.03	0.71
	Post clean-up	99.74	2.06	0.78
Sample 3	Pre clean-up	19.39	1.53	0.59
	Post clean-up	39.78	1.84	0.74
Sample 4	Pre clean-up	10.02	1.68	0.38
	Post clean-up	9.53	1.87	0.60
Sample 5	Pre clean-up	21.95	1.97	0.45
	Post clean-up	14.00	1.89	0.68
Sample 6	Pre clean-up	36.69	1.32	1.09
	Post clean-up	1.04	1.64	0.82
Sample 7	Pre clean-up	9.86	2.18	0.40
	Post clean-up	23.20	1.95	0.30
Sample 8	Pre clean-up	36.52	2.04	0.44
	Post clean-up	34.31	1.85	0.49

As mentioned in the methods, we thus adapted the ABI protocol to enhance the quantity and quality of the RNA, by increasing the time of rotation of the samples as well as increasing the digestion time. The quality was improved with regards to the A260/280 value, however the yield remained low. Table XII (p. 71) shows the results of the extraction of the adapted ABI protocol.

Table XII: Quantity and Quality of RNA extraction from FFPE (Adapted ABI Protocol)

FFPE Tissue	Quantity RNA (ng/µl)	A260/280	A260/230
Sample 1	153.45	1.99	0.68
Sample 2	111.22	1.97	0.96
Sample 3	11.60	1.90	0.09
Sample 4	13.88	1.96	0.09
Sample 5	23.56	1.90	0.83
Sample 6	24.67	1.85	0.67
Sample 7	17.94	2.12	0.07
Sample 8	21.45	2.13	0.08
Sample 9	14.32	2.02	0.09
Sample 10	34.86	1.92	0.51

We then extracted the same RNA from FFPE using both the adapted ABI Protocol and the Roche Protocol and compared the results. Not only were the A260/280 as well as the A260/230 values better, but curves on the Nanodrop were of better quality with less contamination. Overall, the Roche Protocol yielded a better quantity of RNA as well as better quality RNA. These results can be found in Table XIII (p.72).

Table XIII: Quantity and Quality of RNA extraction from FFPE (ABI vs. Roche Protocol)

FFPE Tissue	Protocol	Quantity RNA (ng/µl)	A260/280	A260/230
Sample 1	ABI	99.68	2.07	1.01
	Roche	130.43	1.91	1.70
Sample 2	ABI	140.46	2.00	1.49
	Roche	63.35	2.01	1.89
Sample 3	ABI	42.21	2.03	0.98
	Roche	66.21	1.91	1.38
Sample 4	ABI	37.02	1.99	0.55
	Roche	111.75	1.98	1.69
Sample 5	ABI	50.55	2.03	0.58
	Roche	185.68	2.00	2.06
Sample 6	ABI	35.43	1.86	0.44
	Roche	149.01	2.01	1.83

These results demonstrate that RNA extracted from FFPE tissues as well as fresh tissue is of adequate quality to be used for genetic profiling of tumours. To monitor expression of the selected 25 predictor genes in matched fresh and FFPE tissues, we performed a QPCR using the TaqMan Assays on RNA extracted from each type of sample for the same tumour. This was performed using cDNA from Gene Specific RT. Table XIV (p. 73) describes the quantity and quality of the RNA used for these QPCR experiments.

Table XIV: Quantity and Quality of RNA extraction from FFPE and Fresh Tissue of the same tumour

Tissue	Type tissue	Quantity RNA (ng/µl)	A260/280	A260/230
Sample 1	FFPE	111.75	1.98	1.69
	Fresh	380.64	1.97	1.72
Sample 2	FFPE	149.01	2.01	1.83
	Fresh	348.48	1.99	1.99
Sample 3	FFPE	77.83	1.87	1.59
	Fresh	332.14	1.90	1.94
Sample 4	FFPE	185.68	2.00	2.06
	Fresh	653.65	1.98	2.01

RNA was extracted from Fresh Tissue using CHUM QIAzol Protocol and RNA was extracted from FFPE tissue using the Roche Protocol

Initially we wanted to evaluate if we could achieve a good PCR signal from RNA extracted from FFPE tissues as this RNA, although evaluated to be of good purity, can be quite degraded. Again, each sample was evaluated in a triplicate fashion. All QPCR curves were evaluated and poor curves were eliminated from the final analysis. One sample was randomly selected as the calibrator sample to which other samples were compared. In this case, sample 6 was used as the calibrator sample. Almost all genes had a fold change of greater than 2 or less than 0.5, which is considered significant (Appendix I). The following Heatmap illustrates the expression patterns of the 24 predictive genes on 8 tumour samples where RNA was extracted from FFPE tissues (Figure 10, p. 75), showing an appropriate signal of our genes in downstream QPCR experiments from degraded RNA extracted from FFPE tissues, demonstrating that RNA extracted from FFPE can be used in such QPCR experiments for analysis. A log scale was used to better discriminate variations in gene expression. We note that in this small tumour set, genes associated with a poor response (depicted with a P in

Figure 10, p.75) and genes associated with a good response (depicted with a G in Figure 10, p.75), did not cluster together. It is important to note that the predictive genes selected were meant to idenpendently predict the response to Tamoxifen without necessarily requiring that genes cluster together. In addition, a limitation in this type of analysis is the small sample size and not knowing the patient history associated with these particular tumours and whether or not they responded to Tamoxifen. Nevertheless, this QPCR experiment demonstrated our ability to genetically profile the tumours from RNA extracted from FFPE tissues.

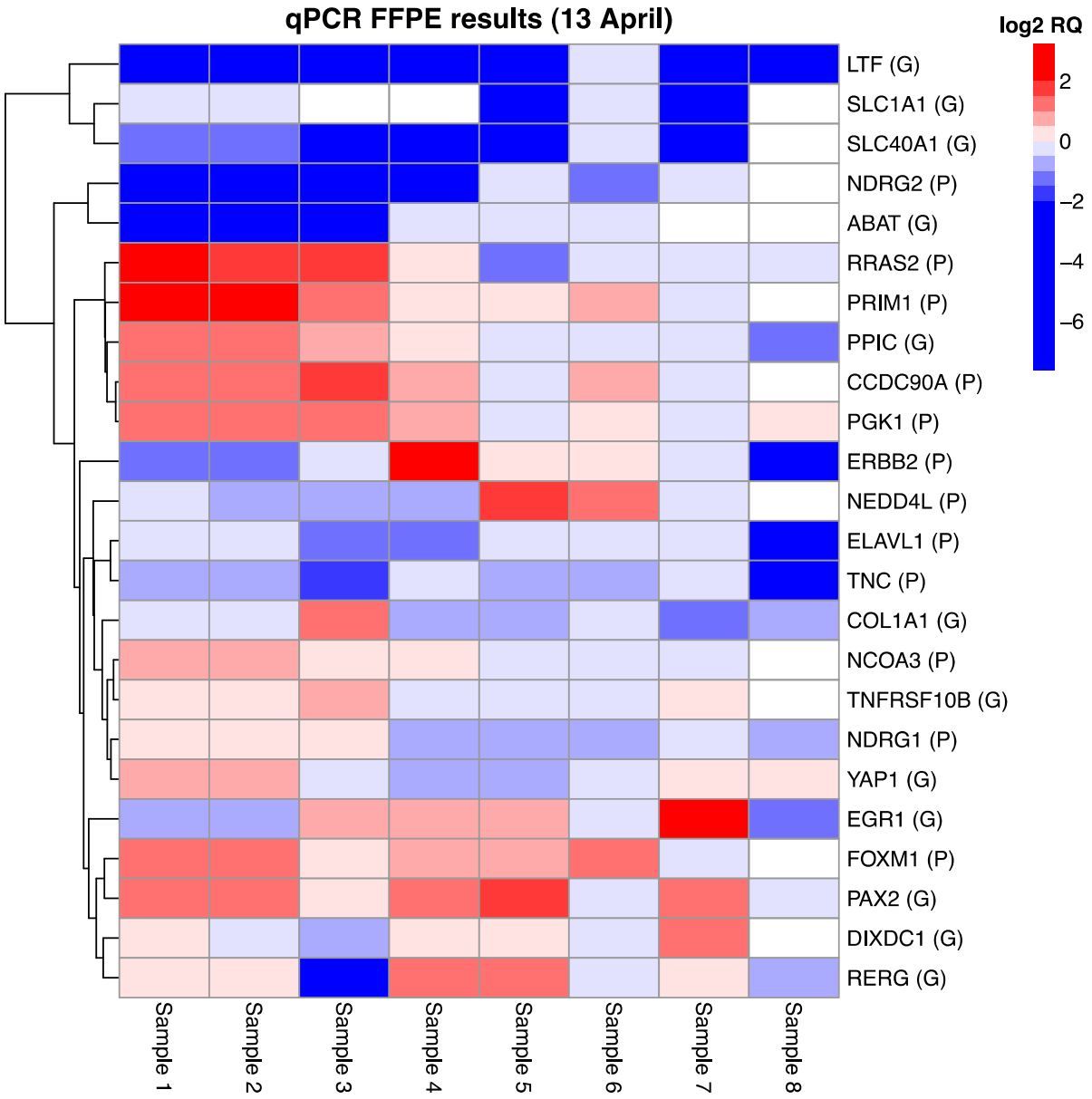


Figure 10: Heatmap results of 24 Predicitve genes in 8 different tumour samples where RNA was extracted from FFPE tissues

Legend: Heatmap representation of the expression of 24 genes in tumours (RNA extracted from FFPE tissues). The level of expression of each gene in each sample is represented in a log-space using a pink-blue color scale Rows of expression matrix, transcripts; columns, profiled samples. Hierarchial clustering (distance: correlation, linkage: ward) was used to regroup genes with similar expression patterns. Genes associated with a good response are depicted with a G. Genes associated with a poor response is depicted with a P.

A QPCR was then performed using the 25 TaqMan assays using the RNA extracted from FFPE and Fresh tissue of the same tumour and the expressions were correlated. Gene-specific reverse transcription was performed on the RNA of these tissues to prepare the cDNA used in the QPCR experiments.

It is important to note that when comparing RNA from FFPE and frozen tissue, the average shift in real-time RT-PCR is in the range of 2-5 Ct. In general, average mRNA fragment size from FFPE is well below 500 bp, and amplicon sizes of quantitative RT-PCR assays should therefore be below 100-120 bp. Of note, the Ct shift between fresh frozen and FFPE derived RNA for a given assay even in the optimal size range (<120 bp) may vary from less than 2 to 5 Ct, although this Ct shift remains constant for a given assay. These differences have to be taken into account especially if RNA transcript levels are directly compared.[126].

Figure 11 (p. 77) demonstrates the distribution of the average Ct between fresh and paired FFPE tumours for both N1 and N2 replicates. The mean difference in average Ct was 1.99, 95% CI (1.60;2.39) which follows the expected Ct shift of 2 to 5 Ct between fresh and FFPE derived RNA. Moreover, this remained constant and all genes were affected in the same way showing a positive correlation between the Fresh and paired FFPE tissues. Paired t-test was performed to assess the significance of the difference in average CT between fresh and FFPE tumours and for both replicates the difference was statistically significant ($p<0.05$).

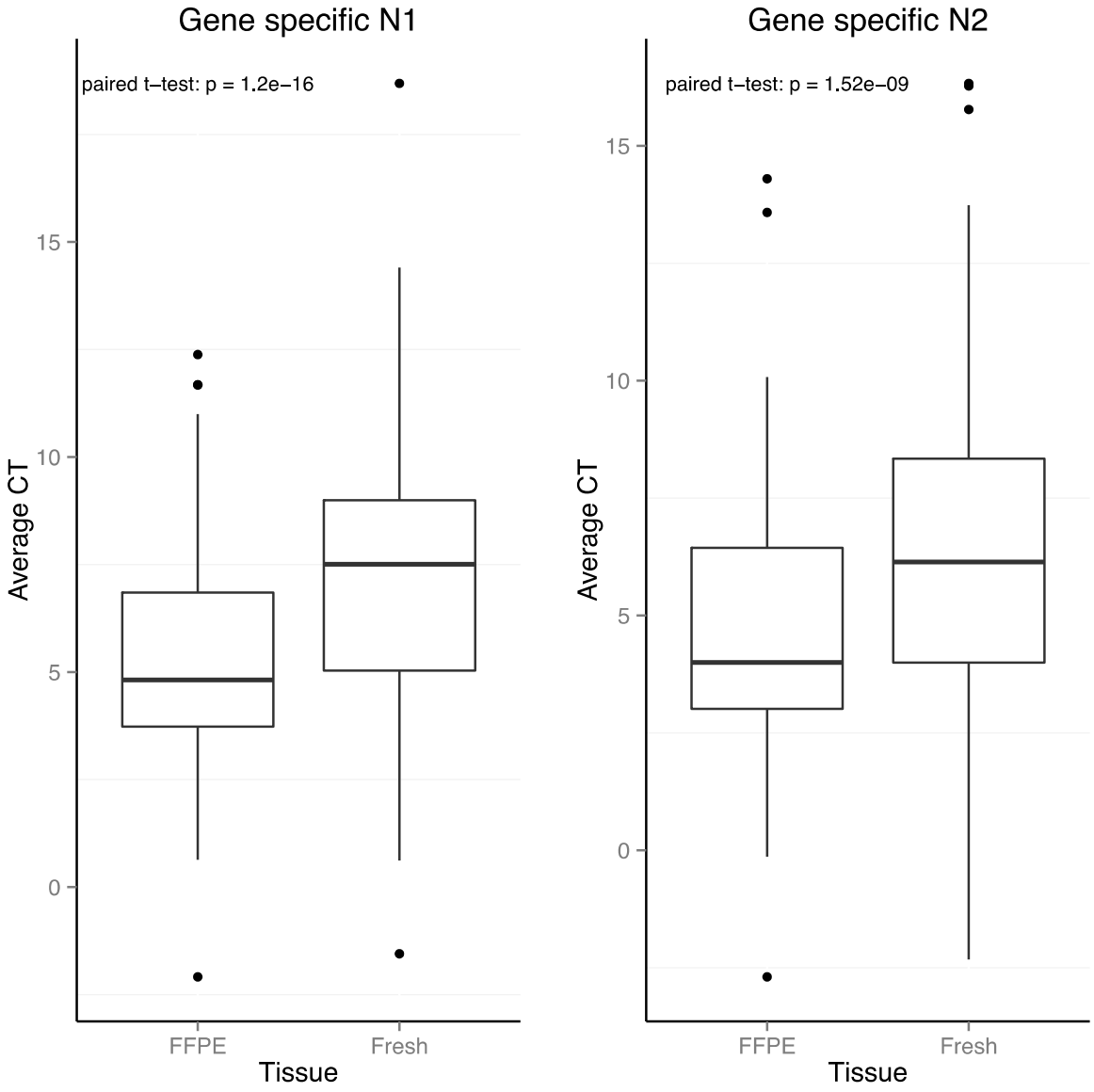


Figure 11: The distribution of the average Ct between fresh and paired FFPE tumours (N1 and N2).

Legend: Boxplot presenting the distribution of average Ct between Fresh and FFPE tumours. (a) Biological replicate 1 and (b) biological replicate 2

The QPCR results of the Fresh and paired FFPE tumour samples were then correlated. Figure 12 (p. 79) demonstrates that when correcting for the systematic difference in RQ between Fresh and FFPE tissues, the Fresh and FFPE samples for the same tumor cluster together despite the degradation of the FFPE RNA. This is a very positive result. When tested statistically, the Pearson correlation (r) between fresh and FFPE was $r = 0.301$ (t-test $p = 0.00232$). Similar results were found in an N2 replicate of this experiment. Not all genes were re-tested in the N2 replicate as there was not enough tumour tissue (RNA) for a complete N2 replicate, however we can see in the genes that were retested, the N2 replicate had a very positive correlation between fresh and FFPE tissues of the same tissue, actually better than the N1 replicate. Figure 13 (p. 80) shows the Pearson correlation between Fresh and FFPE results for both the N1 and N2 replicates.

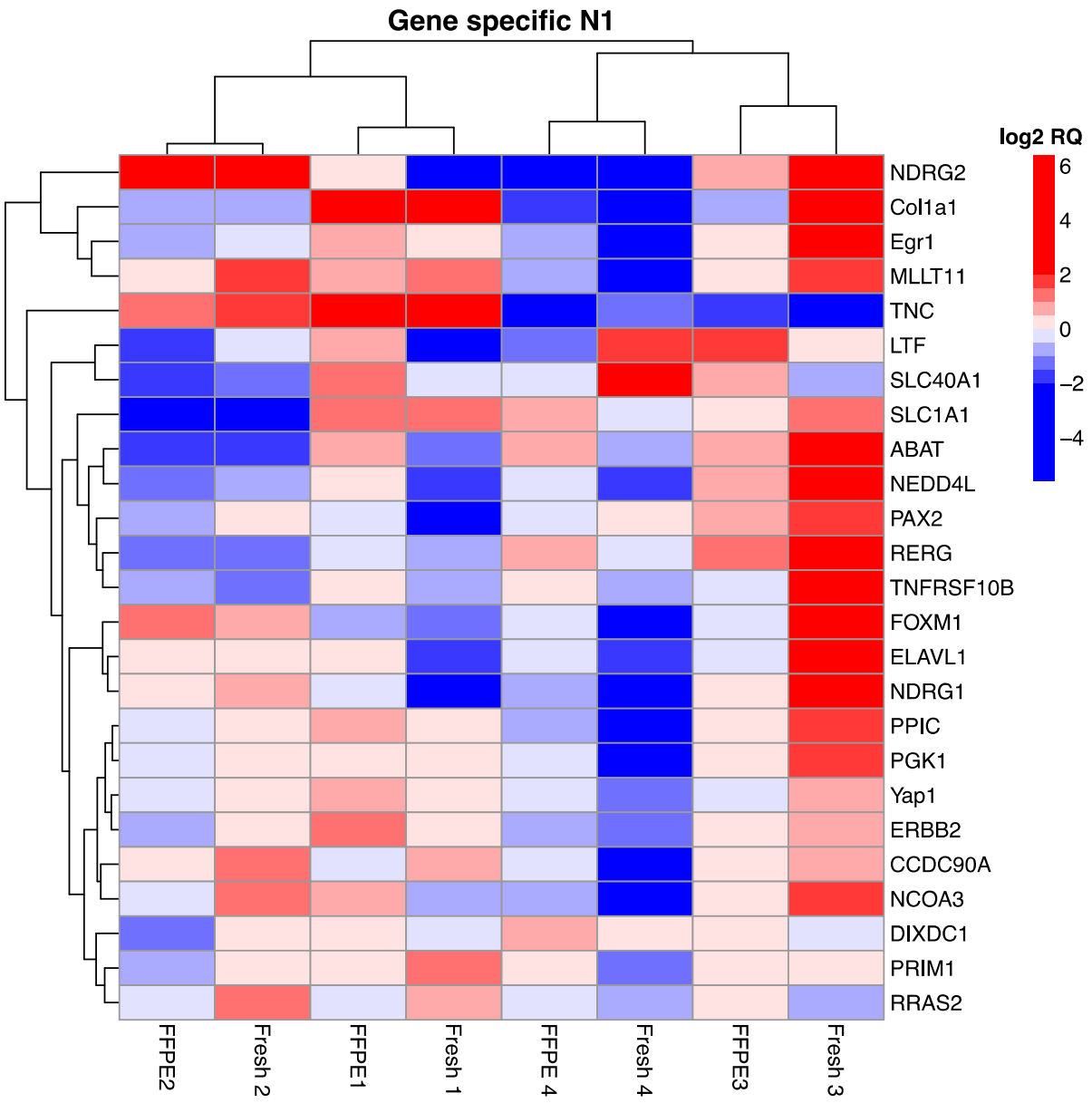


Figure 12: Heatmap correlation of the 25 predictive genes between Fresh and paired FFPE extracted RNA in different tumour samples.

LEGEND: Heatmap representation of the expression of 25 genes in fresh and in FFPE tumours. To correct for the expression bias between fresh and FFPE tumors (FFPE tumors present in average lower RQ) each gene in each tissue type (Fresh or FFPE) was center to 0 prior to sample clustering.The level of expression of each gene in each sample is represented in a log-space using a pink-blue color scale. Rows of expression matrix, transcipts; columns, profiled samples. Hierarchical clustering (distance: correlation, linkage: ward) was used to regroup genes and samples w/ similar expression pattern.

Gene	N1	N2
LTF	-0.31787	-0.17837
RRAS2	-0.25469	0.31452
PRIM1	-0.15111	-0.12939
ELAVL1	-0.13652	NA
TNFRSF10B	-0.12876	-0.0085
SLC40A1	0.121518	NA
FOXM1	0.127796	0.296491
DIXDC1	0.142042	NA
YAP1	0.210701	0.904635
NCOA3	0.288044	0.460766
ERBB2	0.367947	0.50277
ABAT	0.37304	0.426094
COL1A1	0.514235	0.605639
NEDD4L	0.556325	NA
CCDC90A	0.571797	NA
PPIC	0.571797	0.724778
PGK1	0.640434	0.905492
EGR1	0.644288	NA
NDRG2	0.65043	NA
PAX2	0.698446	0.591733
MLLT11	0.756821	NA
RERG	0.820231	NA
SLC1A1	0.863587	NA
NDRG1	0.900493	NA
TNC	0.923141	NA

Figure 13: Pearson correlation between paired Fresh and FFPE tissues (N1 and N2) NA = N2 replicate not done

4.4 Final selection of the gene signature

The LDA card that we had chosen has 24 ports in duplicates. In order to have optimal internal validity we decided to include 4 housekeeping genes to the signature: ACTB, 18S, TBP and PPIA as already discussed. This left 20 predictive genes to choose to include on the final LDA card. Based on the correlation between the genetic expression of the tumours between RNA extracted from FFPE and Fresh Tissue, 5 other genes were eliminated: LTF, PRIM1, RRAS, ELAVL1 and TNFRSF10B. These 5 genes poorly correlated between Fresh and FFPE tissues.

Table XV (p. 82) describes the final 24 genes that were included on the LDA card. There were 4 housekeeping genes, 10 predictors of good outcome and 10 predictors of a poor outcome.

Table XV: The Final 24 genes on the LDA card

Gene	Role
ABAT	Good Predictor
COL1A1	Good Predictor
PAX2	Good Predictor
PPIC	Good Predictor
YAP1	Good Predictor
RERG	Good Predictor
SLC1A1	Good Predictor
DIXDC1	Good Predictor
EGR1	Good Predictor
SLC40A1	Good Predictor
CCDC90A	Poor Predictor
ERBB2	Poor Predictor
FOXM1	Poor Predictor
NCOA3	Poor Predictor
PGK1	Poor Predictor
NDRG1	Poor Predictor
NDRG2	Poor Predictor
TNC	Poor Predictor
NEDD4L	Poor Predictor
MLLT11	Poor Predictor
18S	Housekeeping Gene
TBP	Housekeeping Gene
PPIA	Housekeeping Gene
ACTB	Housekeeping Gene

Discussion and Conclusion

The main purpose of this study was to create a simple PCR tool that could predict the response to Tamoxifen. Since the predictive biomarkers were derived from microarray experiments, we first had to verify the correlation between Q-PCR results and microarray expression profiles .

5.1 Reproducibility between Q-PCR assays and microarray results

Firstly, the predictive genes in our proposed tool were identified using bioinformatics tools based on data sets using the Affymetrix microarray. Although we would expect that QPCR would have similar results to microarray this is not always the case. Our study initially evaluated the signal of our predictive genes by QPCR on 8 well-established breast cancer cell lines and this was correlated to the Affymetrix microarray for the same genes and cell lines. Gene expression was very strongly positively correlated for 25 out of the 30 genes tested. Moreover, the expression patterns obtained were able to discriminate between ER+ and ER- breast cancer cell lines, although with greater accuracy in one replicate versus the other.

5.2 Reproducibility of expression profiles between fresh and FFPE samples

In the clinical setting, breast cancer tissue is often preserved in various forms. In the majority of cases, this is archival formalin-fixed paraffin-embedded tissue. In centers with a tissue bank, tissue may be fresh flash frozen or fresh preserved in RNAlater. For a predictive tool based on gene expression profiles to be used in the clinical setting it must be able to profile genes of a tumour regardless of the way the tumour has been preserved. Thus we first optimized extraction protocols from both fresh frozen and FFPE samples, and then compared expression profiles in different matched fresh/FFPE tumour samples.

With regards to fresh frozen tissue, our study showed that the most crucial step in RNA extraction is ensuring appropriate homogenization of the tissue prior to extraction. This yields both quantity and quality RNA for downstream QPCR experiments. For FFPE tissues, this process is more difficult. The bonds created by the formaldehyde as well as the age of the paraffin blocks play an important role in the quantity and quality of RNA that can be extracted. RNA from FFPE tissues is often difficult to extract and is frequently degraded. Our study has shown that the most crucial part of RNA extraction from FFPE tissues is ensuring longer digestion periods at elevated temperatures, with overnight digestion leading to the best yield of quality and quantity.

Despite the degradation of the RNA, adequate PCR signalling was obtained in our study, and these results were strongly independently positively correlated with fresh tissue of the same tumour, further confirming that RNA from archival paraffin blocks can be reliably used in PCR experiments, if extracted appropriately. Moreover, PCR assays should be kept relatively short in length with an amplicon of less than 120nt and ideally less than 100nt for more reliable results. When the correlation between FFPE and fresh tissues was done, 5 other genes were eliminated due to a negative or poor correlation between the expression of these genes, in fresh tissue versus FFPE of the same tumour. One would expect a similar outcome (upregulation or downregulation) of a gene in a tumour regardless of where this RNA was extracted from (i.e. Fresh or FFPE). In the case of these five genes, however, this was not found. Amplicon size for these genes was adequately small (less than 100nt) and gene specific RT was used to make the cDNA, also signal strength for these genes were good when you look at the fresh and FFPE tissues individually. Despite this, a positive correlation between the expression profile of fresh versus FFPE tissues was not attained. The caveat here is the small

sample size. It is known that although QPCR experiments can be performed using RNA from FFPE tissues, this is not 100% successful. Further experiments with larger sample sizes are needed to confirm these results.

5.3 Tumour clustering based on gene signature expression patterns

Our predictive genes did not cluster in good and poor predictors in different tumours samples. This may be explained by the method in which the genes were selected in our study, as they are independently associated with response to treatment. However, the lack of clustering may be due the small tumour sample with unknown clinical history. Ideally this experiment should be performed again on more tumour blocks.

5.4 Rationale for an LDA platform for this tool

We selected the TaqMan® low-density array (LDA) by Applied Biosystems (ABI) as the platform for our tool. There are multiple advantages to using an LDA card as a platform for gene expression profiling. LDAs are a novel approach to gene expression profiling. Based on real time quantitative-polymerase chain reaction (QRT-PCR), these arrays enable a more focused and sensitive approach to the study of gene expression than gene chips, while offering higher throughput than more established approaches to QRT-PCR, at a lower cost. The main advantage of this tool is that it allows multiple selected genes to be studied from a single sample, and also offers savings in terms of materials. On the 24-gene card that we have selected we can include 20-21 predictive genes with 3-4 housekeeping genes. As ABI has done their own studies of validity on housekeeping genes, all cards have the 18S housekeeping gene occupying one of the ports as part of the tool.

A study by Goulter and colleagues[131] compared LDA's and showed low variability with correlation coefficients close to 1.0. By performing 2-fold and 10-fold serial dilutions of cDNA samples in the LDAs they found a clear linear relationship between the gene expression data points over 5 orders of magnitude. Furthermore, the data generated by the LDA from a cell based pharmacological study were comparable to data generated by conventional QRT-PCR. The other advantage of LDA is its ability to perform a reliable PCR from amounts of cDNA as small as 50-100ng. This is important as RNA extracted from Formalin-fixed paraffin embedded (FFPE)-tissues is often degraded and only small amounts of RNA can be retrieved. Using an LDA as opposed to a standard QPCR method helps optimize our ability to get a reliable expression profile on human tumours requiring very little material.

5.5 Predictive tools in Breast Cancer

The two most common gene signatures used clinically today are the 21-gene recurrence score (Oncotype DX) and the Amsterdam 70-gene prognostic profile (Mammaprint). Oncotype DX is a predictive tool and Mammaprint is a prognostic tool. Multiple studies have shown that although some patients with ER+ tumours derive a benefit from adjuvant chemotherapy, the majority do not. Oncotype DX was created to identify the subset or ER+, node negative patients who would derive benefit and thus avoid over-treating patients who did not. Mammaprint aids as a guide for decision-making with regards to adjuvant therapy.

Oncotype DX

In the initial steps of creating Oncotype DX, a prognostic tool was created to determine which patients with ER+ cancers would recur despite Tamoxifen therapy. In this approach, the

investigators started with the 250 most promising candidate genes selected from the literature. They then used a reverse transcription polymerase chain reaction (RT-PCR)-based method for generating quantitative expression levels of these genes in fixed tissue from 447 patients collected from three largely hormone receptor-positive, node-negative datasets. The result is the recurrence score (RS), which is actually a mathematical formula that includes 16 genes (plus five reference genes) weighted to optimize prediction of distant relapse despite Tamoxifen therapy[132].

The RS was validated in an independent dataset derived from 668 samples (from a total of 2617) collected in the tamoxifen-treated arm of National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14, a prospective randomized clinical trial examining the benefit of adjuvant tamoxifen in hormone receptor-positive, node-negative breast cancer[133].

Once the recurrence score was calculated it was applied to patients that had ER+, node negative cancers that had received Tamoxifen with or without chemotherapy (NSABP B-20 trial)[134], thus predicting those patients that would respond to chemotherapy. In this study, a high RS (greater or equal to 31) predicted benefit of chemotherapy (methotrexate plus fluorouracil with or without cyclophosphamide, (CMF) added to tamoxifen (a decrease in 10-year distant recurrence risk by 28 percent), whereas patients with a low RS derived minimal, if any, benefit from chemotherapy[135]. It is this predictive portion of the tool that is currently being used in practice. Interestingly, in this study they showed that often classical pathologic markers did not correlate with the recurrence scores obtained from the genetic profiling of the tumours. For example, 19% of high-grade tumours had a low RS and 5% of low-grade tumours had a high RS.

Interestingly, Oncotype DX was used to evaluate tumours that were treated with Aromatase Inhibitors and the RS was still accurate in predicting distant recurrence as a prognostic tool, thus demonstrating that the RS can identify relative endocrine insensitivity as a general phenomenon, whether it is tumours treated with Tamoxifen or AI's. It is important to note however, that Oncotype DX with regards to hormone therapy is a prognostic tool as opposed to a predictive one. It is a predictive tool with regards to chemotherapy.

Mammaprint

With regards to Mammaprint, it was developed in a slightly different manner. As already mentioned this is a prognostic tool classifying tumours into a high-risk or low-risk category for recurrence. Investigators from the Netherlands Cancer Institute performed a supervised analysis of gene expression arrays on frozen tissue from primary breast tumours that were used to develop the 70-gene profile[136]. Of the 98 breast tumour samples, 78 (80%) were from node-negative women younger than 55 years, of whom 34 (44%) had developed distant metastasis within five years (and 44 had not). The remaining 20 samples were from women with hereditary breast cancer, BRCA1 mutations, who tend to develop basal-like breast cancers as well as other hereditary features. Supervised analysis of the gene expression arrays selected a 70-gene set with 83% accuracy at differentiating those with distant relapse versus those without.

This was then later validated in a larger study with a cohort of 302 patients. Patients were under age 60, had node-negative T1 to T2 tumours, were treated without adjuvant systemic therapy, and were followed for over 10 years[137]. The 70-gene signature performed independent of clinical variables in predicting time to distant metastasis (hazard ratio, HR

2.13) and overall survival (HR 2.63), but not disease-free survival (HR 1.36). Patients in the gene signature high-risk group had a 10-year overall survival of 70% versus 90% for patients in the gene signature low-risk group[137].

Other genomic tools

Rotterdam 76-gene signature

This gene signature was developed on fresh frozen tissue of 115 tumours that were node negative from women who did not receive adjuvant therapy and were followed for 8 years. Several gene sets were developed, as there were a variety of subtypes including ER- (16 genes) and ER+ tumours (60 genes). It was then validated in an independent set of 180 node-negative, mixed ER+ (84%) and ER- (16%) tumours, showing that with regards to distant metastasis-free survival, the prognostic signature was independent of clinical variables and those with a poor gene signature were 7.41 times more likely to develop distant metastases than those with a good signature. It is important to note however that the ER- group was quite small and thus generalizing to this group is difficult. Moreover, this tool is good for predicting early recurrence, before 5 years but not as good at predicting late recurrences, as is most often seen in patients with ER+ disease[87,138].

Other Genomic Signatures

A number of other signatures are in development and for several types of cancers. With regards to breast cancer, another prognostic tool that is also potentially predictive, is the two-gene signature (HOX13:IL17BR). This was developed using 60 node positive, ER+ tumours. This tool was developed to predict recurrence as well as endocrine therapy sensitivity. The HOXB13 gene is associated with recurrence, while interleukin 17 is associated

with disease free survival[139]. The difficulty with this test is that when performed on fixed tissues it yields mixed results. A refined version of this two-gene signature included a proliferative index, called the molecular grade index (MGI), which could be used in fixed tissue, improving its prognostic capabilities[140].

Finally, the other most promising tool is the "wound response" gene signature. This is based on the core serum response (CSR) genes, which change expression when cultured fibroblasts are activated with serum. This represents the processes of matrix remodelling, cell motility and angiogenesis as seen in wound healing, noted as being important with regards to cancer growth and the tumour microenvironment. Activation of the CSR genes was examined in a 295-patient dataset (used to validate the Amsterdam 70-gene profile). Tumours that expressed the wound response signature as compared to those that did not were found in multivariate analyses to portend a higher likelihood of distant metastasis and poorer overall survival, making it again a prognostic tool[141].

5.6 Clinical Utility of our tool

Although the prognostic gene signatures are helpful in identifying poor prognosis patients and thus may aid in adjuvant treatment selection, so do our traditional pathologic and clinical prognostic factors and these are much cheaper to perform. Prognostic tools do not directly evaluate the benefit of a treatment, and are unable to identify patients in whom treatment may be beneficial or futile, which ultimately may be more clinically useful. Thus predictive tools may be more clinically beneficial. The tools thus far have been developed for ER+ patients, there are very few tools that are meant for ER-, Her-2+ or triple negative patients.

Our tool is a predictive, not a prognostic tool and thus can identify patients that would best respond to Tamoxifen as well identify those most likely to be resistant and may be selected to enter a clinical trial for patients exhibiting hormone therapy resistance. This tool will thus have a direct impact on clinical practice through the development of markers of therapeutic success for treatment with Tamoxifen. Futile treatments would be avoided thus preventing needless side effects, and improved identification of ER+ tumours with a low chance of success to anti-estrogen therapy. This will facilitate research into more appropriate treatments for hormone resistant tumours. There are some suggestions that patients with hormone resistance may be treated with other targeted therapies. In some patients, chronic activation of ER+Her-2+ breast cancer cell lines with heregulin, a ligand for the Her2 receptor family, has shown to lead to down-regulation of ER expression and hormone independence[76]. One report suggests that this effect may be mediated by a co-repressor termed metastasis-associated protein 1 (MTA1) co-repressor and that it can be inhibited by trichostatin A, opening the possibility of pharmacologic reversal of resistance to antiestrogen therapy[142]. These types of discoveries are essential in improving the treatment strategies for ER+ cancers.

Interestingly, Oncotype DX was used to evaluate tumours that were treated with Aromatase Inhibitors and the RS was still accurate in predicting distant recurrence as a prognostic tool, thus demonstrating that the RS can identify relative endocrine insensitivity as a general phenomenon, whether it is tumours treated with Tamoxifen or AI's. These results are promising for our tool as well indicating that our PCR tool may also be applied for patients treated with Aromatase Inhibitors.

5.7 Limitations

Although most of the 30 genes were reliable with a good signal, this was not without limitations. For various reasons 10 genes were eliminated from the final tool. As already mentioned, not all QPCR results correlated with the Affymetrix micro-array and those that poorly correlated were eliminated, as the genes were selected as being predictive based on microarray expression patterns. Moreover, although some genes had a good profile on cell lines, when tested in real human tumours, the QPCR signals were unreliable especially when RNA was extracted from FFPE tissues. Moreover this study was performed on a very small sample size with an uknown clinical history. Perhaps these genes would have had a better signal on different tumours. In order to reliably select genes for future experiments, more tissues should be tested and paired as done in this study with FFPE and Fresh tissue of the same tumour.

Ideally we would have tested ER+ breast cancer tumours that were treated with Tamoxifen with a known clinical history using our LDA tool to confirm that our tool is predictive by creating a recurrence score as was done for OncotypeDx. Unfortunately, there was a lack of tumour samples to test in this manner at the time of this project.

5.8 Future Perspectives

The purpose of this study was to develop a platform for predictive biomarkers absed on gene expression pattercreate the actual tool. Validity of the tool with tumour samples will be performed as a future study. Although it would have been ideal to extract the RNA from archival samples with a known clinical history, at the current time, there was unexpectedly a

shortage of this tissue in order to perform the actual experiment. Archival tissues are being sought and identified in order to complete this portion of the project.

Patient Selection and History

Tissues will be obtained from the CHUM-Hôtel-Dieu hospital breast cancer tumor bank, from patients that have been followed in two prospective NSABP trials. We will be using tissues from women in the NSABP-B-14 protocol, which compared patients receiving Tamoxifen monotherapy vs Placebo and the NSABP-B-27 protocol where each patient received Tamoxifen therapy and various chemotherapy combinations. We have access to the recorded pathological and clinical data (5-10 years follow-up) of these patients allowing us to know the eventual clinical outcome of these patients. The selection of tumors will include samples from 30-40 patients with ER-positive tumors who received Tamoxifen therapy and have recurred or died after 5-10 years of follow-up versus 30-40 patients with ER-positive tumours who received Tamoxifen therapy and show no evidence of disease after 5-10 years of follow-up. This will allow us to distinguish the molecular profile of a cancer that responds versus one that does not.

The complete consent form and research protocol was written up (Appendix L) and was approved by both the Scientific (Appendix M) and Ethics Committees (Appendix N) of the Hôtel-Dieu Hospital of the CHUM.

Creation of Recurrence Score

Once the gene profiles of patients that have recurred is compared to those that have ot, a recurrence score (mathematical algorithm) can be calculated for use as a guide to clinical

practice for the use of hormone therapy in patients with ER+ tumours. This is a future perspective once the RNA from archival tissue is assessed with our LDA tool.

6. References

[1] Canadian Cancer Society/National Cancer Institute of Canada. Canadian Cancer Statistics 2010, Toronto, Canada, 2010.

[2] American Cancer Society Breast Cancer Facts and Figures 2008.

[3] Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005; 55:74-108.

[4] Jatoi I, Chen BE, Anderson WF, Rosenberg PS. Breast cancer mortality trends in the United States according to estrogen receptor status and age at diagnosis. *J Clin Oncol* 2007; 25:1683-90.

[5] Data on SEER (Surveillance, Epidemiology and End Results) cancer statistics.

[6] Degnim AC, Visscher DW, Berman HK et al. Stratification of breast cancer risk in women with atypia: a Mayo cohort study. *J Clin Oncol* 2007; 25:2671-7.

[7] Fisher B, Dignam J, Wolmark N et al. Tamoxifen in the treatment of intraductal breast cancer: National Surgical Adjuvant Breast and Bowel Project B-24 randomised controlled trial. *Lancet* 1999; 353:1993-2000.

[8] Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001; 358:1389-99.

[9] Kelsey, JL, Gammon, MD, John, EM. Reproductive factors and breast cancer. *Epidemiol Rev* 1993; 15:36.

[10] Brinton LA, Schairer C, Hoover RN et al. Menstrual factors and risk of breast cancer. *Cancer Invest* 1988; 6:245-54.

[11] Missmer SA, Eliassen AH, Barbieri RL et al. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J Natl Cancer Inst* 2004; 96:1856-65.

[12] McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006; 15:1159-69.

[13] Vachon CM, Brandt KR, Ghosh K et al. Mammographic breast density as a general marker of breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2007; 16: 43–49.

[14] Boyd NF, Lockwood GA, Martin LJ et al. Mammographic densities and breast cancer risk. Breast Dis 1998; 10: 113–126.

[15] Chen WY, Hankinson SE, Schnitt SJ et al. Association of hormone replacement therapy to estrogen and progesterone receptor status in invasive breast carcinoma. *Cancer* 2004; 101:1490-500.

[16] Calle EE, Feigelson HS, Hildebrand JS et al. Postmenopausal hormone use and breast cancer associations differ by hormone regimen and histologic subtype. *Cancer* 2009; 115:936-45.

[17] Bhatia S, Robison LL, Oberlin O et al. Breast cancer and other second neoplasms after childhood Hodgkin's disease. *N Engl J Med* 1996; 334:745-51.

[18] Van den Brandt PA, Spiegelman D, Yaun SS et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol* 2000; 152:514-27.

[19] Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. *JAMA* 2001; 286:2143-51.

[20] Allen NE, Beral V, Casabonne D et al. Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst* 2009; 101:296-305.

[21] Perou CM, Sørlie T, Eisen MB et al. Molecular portraits of human breast tumours. *Nature* 2000; 406:747.

[22] Prat A, Parker JS, Karginova O et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 2010; 12:R68.

[23] Sørlie T, Perou C, Tibshirani R et al. Gene Expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* 2001; 98:10869-10874.

[24] Slamon D, Eiermann W, Robert N et al. Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med* 2011; 365:1273.

[25] Sorlie T, Tibshirani R, Parker J et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 100:8418.

[26] Radisky DC. Epithelial-Mesenchymal Transition. *J cell sci* 2005; 118:4325-26.

[27] Deroo BJ, Korach KS. Estrogen receptors and human disease. *J Clin Invest* 2006; 116:561.

[28] Mangelsdorf DJ, Thummel C, Beato M, et al. The nuclear receptor superfamily: the second decade. *Cell* 1995; 83:835.

[29] Acevedo ML, Kraus WL. Transcriptional activation by nuclear receptors. *Essays Biochem* 2004; 40:73.

[30] Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci U S A* 1996; 93:5925.

[31] Tsai MJ, O'Malley BW. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu Rev Biochem* 1994; 63:451.

[32] Glass CK. Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr Rev* 1994; 15:391.

[33] Klinge CM. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res* 2001; 29:2905.

[34] Metzger D, Ali S, Bornert JM, Chambon P. Characterization of the amino-terminal transcriptional activation function of the human estrogen receptor in animal and yeast cells. *J Biol Chem* 1995; 270:9535.

[35] Edwards DP. The role of coactivators and corepressors in the biology and mechanism of action of steroid hormone receptors. *J Mammary Gland Biol Neoplasia* 2000; 5:307.

[36] McEwan IJ. Gene regulation through chromatin remodelling by members of the nuclear receptor superfamily. *Biochem Soc Trans* 2000; 28 **:**369-373.

[37] McKenna NJ, Lanz RB, O'Malley BW. Nuclear receptor coregulators : cellular and molecular biology. *Endocr Rev* 1999 ; 20 :321-44.

[38] Kraus WL, Kadonaga JT. p300 and estrogen receptor cooperatively activate transcription via differential enhancement of initiation and reinitiation. *Genes Dev* 1998 ;12 :331-42.

[39] Ring A, Dowsett M. Mechanisms of tamoxifen resistance. *Endo-Related Cancer* 2004; 11: 643-658.

[40] Ignar-Trowbridge DM, Nelson KG, Bidwell MC, Curtis SW, Washburn TF, McLachlan JA, Korach KS. Coupling of dual signaling pathways: epidermal growth factor action involves the estrogen receptor. *Proc Natl Acad Sci U S A* 1992; 89:4658.

[41] Stoica A, Saceda M, Fakhro A, et al. Role of insulin-like growth factor-I in regulating estrogen receptor-alpha gene expression. *J Cell Biochem* 2000; 76:605.

[42] Yee D, Lee AV. Crosstalk between the insulin-like growth factors and estrogens in breast cancer. *J Mammary Gland Biol Neoplasia* 2000; 5:107.

[43] Bocchinfuso WP, Lindzey JK, Hewitt SC, et al. Induction of mammary gland development in estrogen receptor-alpha knockout mice. *Endocrinology* 2000; 141:2982.

[44] Couse JF, Hewitt SC, Bunch DO, Sar M, Walker VR, Davis BJ, Korach KS. Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. *Science* 1999; 286:2328.

[45] Love RR, Mazess RB, Barden HS et al. Effects of tamoxifen on bone mineral density in postmenopausal women with breast cancer. *N Engl J Med* 1992; 326:852-6.

[46] Grey AB, Stapleton JP, Evans MC, Reid IR. The effect of the anti-estrogen tamoxifen on cardiovascular risk factors in normal postmenopausal women. *J Clin Endocrinol Metab* 1995; 80:3191.

[47] Fisher B, Costantino JP, Wickerham DL, Cecchini RS, Cronin WM, Robidoux A, Bevers TB et al. Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst* 2005; 97:1652.

[48] Osborne CK. Tamoxifen in the treatment of breast cancer. *N Engl J Med* 1998; 339:1609.

[49] Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* 2011; 378:771.

[50] Beato M, Sánchez-Pacheco A. Interaction of steroid hormone receptors with the transcription initiation complex. *Endocr Rev* 1996; 17:587.

[51] Beato M, Candau R, Chavez S, Mows C, Truss M. Interaction of steroid hormone receptors with transcription factors involves chromatin remodelling. *J Steroid Biochem Mol Biol* 1996; 56:47.

[52] Smith IE, Dowsett M. Aromatase inhibitors in breast cancer. *N Engl J Med* 2003; 348:2431.

[53] Miller WR. Biological rationale for endocrine therapy in breast cancer. *Best Pract Res Clin Endocrinol Metab* 2004; 18:1.

[54] McKeage K, Curran MP, Plosker GL. Fulvestrant: a review of its use in hormone receptor-positive metastatic breast cancer in postmenopausal women with disease progression following antiestrogen therapy. *Drugs* 2004; 64:633-48.

[55] Jones SE. Fulvestrant: an estrogen receptor antagonist that downregulates the estrogen receptor. *Semin Oncol* 2003; 30(Suppl 16):14-20.

[56] Johnston SJ, Cheung KL. Fulvestrant – a novel endocrine therapy for breast cancer. *Curr Med Chem* 2010; 17:902-14.

[57] Dowsett M, Cuzick J, Ingle J, Coates A, Forbes J, Bliss J, Buyse M. Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen. *J Clin Oncol* 2010; 28:509.

[58] Jin H, Tu D, Zhao N, Shepherd LE, Goss PE. Longer-term outcomes of letrozole versus placebo after 5 years of tamoxifen in the NCIC CTG MA.17 trial: analyses adjusting for treatment crossover. *J Clin Oncol* 2012; 30:718.

[59] Jordan VC. Selective estrogen receptor modulation: concept and consequences in cancer. *Cancer Cell* 2004; 5:207.

[60] Smith CL, Nawaz Z, O'Malley BW. Coactivator and corepressor regulation of the agonist/antagonist activity of the mixed antiestrogen, 4-hydroxytamoxifen. *Mol Endocrinol* 1997; 11:657.

[61] McInerney EM, Katzenellenbogen BS. Different regions in activation function-1 of the human estrogen receptor required for antiestrogen- and estradiol-dependent transcription activation. *J Biol Chem* 1996; 271:24172-8.

[62] Jackson TA, Richer JK, Bain DL, et al. The partial agonist activity of antagonist-occupied steroid receptors is controlled by a novel hinge domain-binding coactivator L7/SPA and the corepressors N-CoR or SMRT. *Mol Endocrinol* 1997; 11:693.

[63] Arpino G, Weiss H, Lee AV, et al. Estrogen receptor-positive, progesterone receptor-negative breast cancer: association with growth factor receptor expression and tamoxifen resistance. *J Natl Cancer Inst* 2005; 97:1254.

[64] Nicholson RI, Gee JM. Oestrogen and growth factor cross-talk and endocrine insensitivity and acquired resistance in breast cancer. *Br J Cancer* 2000; 82:501.

[65] Osborne CK, Bardou V, Hopp TA, et al. Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst* 2003; 95:353.

[66] Hosey AM, Gorski JJ, Murray MM, et al. Molecular basis for estrogen receptor alpha deficiency in BRCA1-linked breast cancer. *J Natl Cancer Inst* 2007; 99:1683.

[67] Sun D, Sharma AK, Dellinger RW, et al. Glucuronidation of active tamoxifen metabolites by the human UDP glucuronosyltransferases. *Drug Metab Dispos* 2007; 35:2006.

[68] Desta Z, Ward BA, Soukhova NV, Flockhart DA. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *J Pharmacol Exp Ther* 2004; 310:1062.

[69] Sestak I, Kealy R, Nikoloff M, et al. Relationships between CYP2D6 phenotype, breast cancer and hot flushes in women at high risk of breast cancer receiving prophylactic tamoxifen: results from the IBIS-I trial. *Br J Cancer* 2012; 107:230.

[70] Goetz MP, Rae JM, Suman VJ, et al. Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J Clin Oncol* 2005; 23:9312.

[71] Regan MM, Leyland-Jones B, Bouzyk M, et al. CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the breast international group 1-98 trial. *J Natl Cancer Inst* 2012; 104:441.

[72] Rae JM, Drury S, Hayes DF, et al. CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *J Natl Cancer Inst* 2012; 104:452.

[73] Gottardis MM, Jordan VC. Development of tamoxifen-stimulated growth of MCF-7 tumors in athymic mice after long-term antiestrogen administration. *Cancer Res* 1988; 48:5183.

[74] Lavinsky RM, Jepsen K, Heinzel T, et al. Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc Natl Acad Sci U S A* 1998; 95:2920.

[75] El-Ashry D, Miller DL, Kharbanda S, et al. Constitutive Raf-1 kinase activity in breast cancer cells induces both estrogen-independent growth and apoptosis. *Oncogene* 1997; 15:423.

[76] Pietras RJ, Arboleda J, Reese DM, et al. HER-2 tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth in human breast cancer cells. *Oncogene* 1995; 10:2435.

[77] Holm C, Rayala S, Jirström K, et al. Association between Pak1 expression and subcellular localization and tamoxifen resistance in breast cancer patients. *J Natl Cancer Inst* 2006; 98:671.

[78] Mariadason JM, Augenlicht LH, Arango D. Microarray analysis in the clinical management of cancer. *Hematol Oncol Clin North Am* 2003; 17:377.

[79] Ma, PC. Overview of gene expression profiling, proteomics, and microRNA profiling in clinical oncology. Available at : http://www.uptodate.com (Accessed March 25, 2013).

[80] Sinicropi D, Cronin M, Liu Ml-L. (2007) BioMEMS and Biomedical Nanotechnology : Volume II : Micro/Nano Technologies for Genomics and Proteomics. (Gene Expression Profiling utilizing Microarray technology and RT-PCR). California : Springer.

[81] Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351:2817.

[82] van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415:530.

[83] Boudreau V, Deschênes J, Laperrière D, Aid M, White JH, Mader S. Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nuc. Acid Res* 2008; 36:76-93.

[84] Oh DS, Troester MA, Usary J et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006; 24:1656-1664.

[85] Yu J, Yu J, Cordero KE et al. A transcriptional fingerprint of estrogen in breast cancer predicts patient survival. *Neoplasia* 2008; 10:79-88.

[86] Loi S, Haibe-Kains B, Desmedt C et al. Predicting prognosis using molecular profiling in estrogen receptor positive breast cancer treated with tamoxifen. *BMC Genomics* 2008; 9:239.

[87] Desmedt C, Piette F, Loi S et al. Strong time dependence of 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 2007; 13:3207-14.

[88] Sotiriou C, Wirapati P, Loi S. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006; 98: 262-72.

[89] Bezault J, Bhimani R, Wiprovnick J, Furmanski P. Human lactoferrin inhibits growth of solid tumors and development of experimental metastases in mice. *Cancer Res* 1994. 54: 2310–2312.

[90] Vecchi M, Confalonieri S, Nuciforo P et al. Breast cancer metastases are molecularly distinct from their primary tumors. *Oncogene* 2008; 27:2148-2158.

[91] Ganten TM, Sykora J, Koschny R et al. Prognostic significance of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) receptor expression in patients with breast cancer. *J Mol Med* 2009; 79:995-1007.

[92] Yuan M, Tomlinson V, Lara R et al. Yes-associated protein (YAP) functions as a tumor suppressor in breast. *Cell Death and Differentiation* 2008 ; 15 :1752-1759.

[93] Kerr II DA, Wittliff JL. A Five-Gene Model Predicts Clinical Outcome in ER+/PR+, Early-Stage Breast Cancers Treated with Adjuvant Tamoxifen. *Horm Canc* 2011; 2:261-271.

[94] Finlin BS, Gau C-L, Murphy GA et al. RERG is a novel ras-related, estrogen regulated and growth-inhibitory gene in breast cancer. *Jour Biol Chem* 2001; 276:42259-42267.

[95] Habashy HO, Powe DG, Glaab E et al. RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: a marker of ER-positive luminal-like subtype). *Breast Cancer Res Treat* 2011; 128:315-326.

[96] Beauchemin D, Lacombe C, Van Themsche C. PAX2 is activated by estradiol in breast cancer cells of the luminal subgroup selectively, to confer a low invasive phenotype. *Molecular Cancer* 2011; 10:148-58.

[97] Liu Q, Li J-g, Zheng X-y, Jin F, Dong H-t. Expression of CD133, PAX2, ESA, and GPR30 in invasive ductal breast carcinomas. *Chinese Medical Journal* 2009;122:2763-2769.

[98] Hurtado A, Holmes KA, Geistlinger TR et al. *ERBB2* regulation by Estrogen Receptor-Pax2 determines tamoxifen response. *Nature* 2008; 456: 663–666.

[99] Jiang XP, Elliott RL, Head JF. Manipulation of Iron Transporter Genes Results in the Suppression of Human and Mouse Mammary Adenocarcinomas. *Anticancer Research* 2010; 30: 759-766.

[100] Miller LD, Coffman LG, Chou JW et al. An Iron Regulatory Gene Signature Predicts Outcome in Breast Cancer. *Cancer Res* 2011 ; 71:6728-6737.

[101] Choi YJ, Choi YL, Cho EY et al. Expression of BMI-1 protein in tumor tissue is associated with favorable prognosis in breast cancer patients. *Breast Cancer Res Treat* 2009 ; 113 :83-93.

[102] Wang Y, Zhe H, Ding Z, Gao P, Zhang N, Li G. Cancer stem cell marker Bmi-1 Expression is associated with Basal-like Phenotype and Poor Survival in Breast Cancer. *World J Surg* 2012 ; 36 : 1189-1194.

[103] Zhou W, Guan X, Wang L, Liao Y, Huang J. p12CDK2-AP1 inhibits breast cancer cell proliferation and in vivo tumor growth. *J Cancer Res Clin Oncol* 2012; 138:2085–2093.

[104] Flageng MH, Haugen Moi LL, Dixon JM et al. Nuclear receptor co-activators and HER-2/neu are upregulated in breast cancer patients during neo-adjuvant treatment with aromatase inhibitors. *Br J Cancer* 2009 :1-8.

[105] Carr JR, Kiefer MM, Park HJ et al. FoxM1 regulates mammary luminal cell fate. *Cell Rep* 2012; 1:715-729.

[106] Millour J, Constantinidou D, Stavropoulou AV et al. FOXM1 is a transcriptional target of ERα and has a critical role in breast cancer endocrine sensitivity and resistance. *Oncogene* 2010; 29:2983-2995.

[107] Zhao C, Yasui K, Lee CJ, Kurioka H, Hosokawa Y, Oka T, Inazawa J. Elevated expression levels of NCOA3, TOP1, and TFAP2C in breast tumors as predictors of poor prognosis. *Cancer* 2003; 98:18-23.

[108] Su Q, Hu S, Gao H et al. Role of AIB1 for Tamoxifen resistance in estrogen receptor-positive breast cancer cells. *Oncology* 2008; 75:159-168.

[109] Lahusen T, Henke RT, Kagan BL, Wellstein A, Riegel AT. The role and regulation of the nuclear receptor co-activator AIB1 in breast cancer. *Breast Cancer Res Treat* 2009; 116:225-237.

[110] Zhang DH, Tai LK, Wong LL, Chiu LL, Sethi SK, Koay ESC. Proteomic study reveals that proteins involved in metabolic and detoxification pathways are highly expressed in HER-2/neu-positive breast cancer. *Molecular & Cellular Proteomics* 2005; 4:1686 –1696.

[111] Mendes-Pereira AM, Sims D, Dexter T et al. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to Tamoxifen. *PNAS* 2012; 109:2730-2735.

[112] Rokavec M, Schroth W, Amaral SMC et al. A polymorphism in the TC21 promoter associates with an unfavorable Tamoxifen treatment outcome in breast cancer. *Cancer Res* 2008; 68:9799–808.

[113] Guttery DS, Shaw JA, Loyd K, Pringle JH, Walker RA. Expression of tenascin-C and is isoforms in the breast. *Cancer Metastasis Rev* 2010; 29:595-606.

[114] Helleman J, Jansen MPHM, Ruigrok-Ritstier K et al. Association of an extracellular matrix gene cluster with breast cancer prognosis and endocrine therapy response. *Clin Cancer Res* 2008; 14:5555-5564.

[115] Oskarsson T, Acharyya S, Zhang XHF et al. Breast cancer cels produce tenascin C as a metastatic niche component to colonize the lungs. *Nature Med* 2011; 17:867-875.

[116] Nagaharu K, Zhang X, Yoshida T et al. Tanscin C induces epithelial-mesenchymal transition-like change accompanied by SRC activation and focal adhesion kinase phosphorylation in human breast cancer cells. *Amer J Path* 2011; 178:754-763.

[117] Nagai MA, Gerhard R, Fregani JHTG et al. Prognostic value of NDRG1 and SPARC protein expression in breast cancer patients. *Breast Cancer Res Treat* 2011; 126:1–14.

[118] Chang X-Z, Li D-Q, Hou Y-F et al. Identification of the functional role of AF1Q in the progression of breast cancer. *Breast Cancer Res Treat* 2008; 111:65–78.

[119] Oishi Y, Nagasaki K, Miyata S et al. Functional pathway characterized by gene expression analysis of supraclavicular lymph node metastasis-positive breast cancer. *J Hum Genet* 2007; 52:271–279.

[120] Miller LD, Smeds J, George J et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 2005; 102:13550-5.

[121] Chin K, DeVries S, Fridlyand J et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 2006; 10:529-41.

[122] Loi S, Haibe-Kains B, Desmedt C et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007; 25:1239-46.

[123] Zhang Y, Sieuwerts AM, McGreevy M et al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* 2009; 116:303-9.

[124] Fend F, Hewitt S, Jaggi R, Knudsen B, Luo J, Ribeiro-Silva A. FFPE and RNA extraction: A troubleshooting guide. Genome Technology 2009; 5-14.

[125] Abramovitz M, Ordanic-Kodani M, Wang Y et al. Optimization of RNA extraction from FFPE tissues for expression profiling in the DASL assay. *Biotechniques* 2008; 44: 417–423.

[126] Chung J-Y, Braunschweig T, Hewitt SM. Optimization of recovery of RNA from formalin-fixed paraffin-embedded tissue. *Diagn Mol Pathol* 2006; 15:229–236.

[127] Jaggi R, Quabius S. New Technologies for FFPE Samples improved RNA Isolation and Novel cDNA priming for qPCR and for Universal mRNA Amplification. *J Biomol Tech* 201 ; 21 : S43.

[128] Al Muktafi S, Wang, D-Y, Youngson BJ, Miller N, Boerner S, Done SJ, Wey LL. Clinical relevance of DNA microarray analyses using archival formalin-fixed paraffin-embedded breast cancer specimens. *BMC Cancer* 2011; 11:253-266.

[129] Quality Assessment of Total RNA. http://biomedicalgenomics.org/index.html (Accessed November 13, 2012).

[130] Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 2002; 3:34.1-34.11.

[131] Goulter AB, Harmer DW, Clark KL. Evaluation of low density array technology for quatitative parallel measurement of multiple genes in human tissue. *BMC Genomics* 2006; 7 :1-10.

[132] Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351:2817.

[133] Fisher B, Costantino J, Redmond C, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med* 1989; 320:479-84.

[134] Fisher B, Dignam J, Wolmark N, et al.: Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 1997; 89:1673-82.

[135] Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor- positive breast cancer. *J Clin Oncol* 2006; 24:3726.

[136] van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415:530.

[137] Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006; 98:1183.

[138] Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005; 365:671.

[139] Ma XJ, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004; 5:607.

[140] Ma XJ, Salunga R, Dahiya S, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res* 2008; 14:2601.

[141] Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc *Natl Acad Sci U S A* 2005; 102:3738.

[142] Mazumdar A, Wang RA, Mishra SK, et al. Transcriptional repression of oestrogen receptor by metastasis-associated protein 1 corepressor. *Nat Cell Biol* 2001; 3:30.

7. Appendices

APPENDIX A. **Isolation of RNA using Trizol Reagent**

1. Cells Grown in Monolayer

- Lyse cells directly in a culture dish by adding 1 ml of TRIzol® Reagent *(Invitrogen Cat. No. 15596-018)* to a 3.5 cm diameter dish, and passing the cell lysate several times through a pipette. The amount of TRIzol® Reagent added is based on the area of the culture dish (1 ml per 10 cm2) and not on the number of cells present. An insufficient amount of TRIzol® Reagent may result in contamination of the isolated RNA with DNA.

2. Phase Separation

- Incubate the homogenized samples for 5 minutes at 15-30°C to permit the complete dissociation of nucleoprotein complexes.
- Add 0.2 ml of chloroform per 1 ml of TRIzol® Reagent
- Cap sample tubes securely
- Shake tubes vigorously by hand for 15 seconds
- Incubate them at 15-30°C for 2-3 minutes
- Centrifuge the samples at no more than 12,000 x g for 15 minutes at 2-8°C
- Following centrifugation, the mixture separates into a lower red, phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The volume of the aqueous phase is about 60% of the volume of the TRIzol® Reagent used for homogenization.

3. RNA Precipitation

- Transfer the aqueous phase to a fresh tube
- Precipitate the RNA from the aqueous phase by mixing with isopropyl alcohol. Use 0.5 ml of isopropyl alcohol per 1 ml of TRIzol® Reagent used for the initial homogenization.
- Incubate samples at 15-30°C for 10 minutes
- Centrifuge at no more than 12,000 x g for 10 minutes at 2-8°C. The RNA precipitate, often visible before centrifugation, forms a gel-like pellet on the side and bottom of the tube.

4. RNA Wash

- Remove the supernatant
- Wash the RNA pellet once with 75% ethanol, adding at least 1 ml of 75% ethanol per 1 ml of TRIzol® Reagent used for the initial homogenization.
- Mix the sample by vortexing
- Centrifuge at no more than 7500 x g for 5 minutes at 2-8°C.

5. Redissolving the RNA

- At the end of the procedure, briefly dry the RNA pellet (air-dry for 5-10 minutes). It is important not to let the RNA pellet dry completely as this will greatly decrease its solubility.
- Dissolve RNA in RNAse-free water by passing the solution a few times through a pipette tip
- Incubate for 10 minutes at 55-60°C, store at -70°C.

APPENDIX B. **ABI Protocol – RNA Extraction from Fresh tissue**

1. **Liquid nitrogen pre-processing**
   - Remove the tissue sample from storage and place on dry ice.
   - Place the amount of tissue sample to be processed (0.1 to 3 g) in a pre-cooled (–80 °C), clean mortar.
   - Freeze the tissue sample thoroughly by adding a small volume (5 to 20 mL) of liquid nitrogen.
   - Allow most of the liquid nitrogen to evaporate, leaving ≥1 to 2 mL.
   - Grind the tissue with a pre-cooled (–80 °C) pestle until a fine powder is obtained.
   - IMPORTANT! Make sure the tissue sample remains frozen at all times by adding more liquid nitrogen if necessary
   - Aliquot the greater of 200 µLor 10–12 µL/mg of tissue Lysis/Binding Solution into a wide-mouth container.
   - Using a pre-cooled (–80 °C) metal spatula/scoop, transfer the powder to a pre-weighed and cleaned sterile 50-mL tube and into the premeasured Lysis/Binding Solution, then mix rapidly.

2. **Collect the cells and remove the culture medium**
   - *Suspension cells*: pellet the cells at low speed, and discard the culture medium.
   - *Adherent cells*: aspirate and discard the culture medium from the culture vessel or trypsinize cells to detach them from the growing surface

3. **Lysis and Binding**
   - Add 200–700 µL Lysis/Binding Solution to 100–107 cells and vortex or pipette the lysate up and down several times to completely lyse the cells and to obtain a homogenous lysate. Cells will lyse immediately upon exposure to the solution. Use the low end of the volume range (~200 µL) of Lysis/Binding Solution for small numbers of cells (hundreds), and use closer to 700 µL when isolating RNA from larger numbers of cells (millions).

4. **RNA Isolation**
   - Heat an aliquot of Elution Solution (typically ~50–200 µL per prep) in an RNase-free microcentrifuge tube in a heat block set to 70–80°C.
   - Ensure that filter cartridges are adequate with glass fiber filter down at the bottom of the cartridge
   - Lysate should be somewhat viscous but not too viscous. If too viscous homogenize further or add more lysis/binding solution and homogenize further
   - Centrifuge at top speed for 2-3 minutes to remove debris (optional)
   - Add an equal volume of 64% Ethanol to the lysate and mix gently but thoroughly by carefully pipetting or vortexing, or by inverting the tube several times.

- Apply the lysate/ethanol mixture (from the previous step) to a Filter Cartridge assembled in a Collection Tube. The maximum volume that can be applied at one time is ~700 µL.
- Centrifuge at RCF10,000–15,000 x g (typically10,000–14,000rpm) for ~15 sec–1 min or until the lysate/ethanol mixture is through the filter.
- Discard the flow-through and reuse the Collection Tube for the washing steps.
- Repeat as necessary with ~700 µL aliquots until all of the sample has been drawn though the filter. Generally up to ~2 mL of sample mixture can be passed through the filter without clogging or exceeding its RNA binding capacity.
- Apply 700 µL Wash Solution #1 to the Filter Cartridge.
- Draw the washes through the filter as in the previous step. Discard the flow-through and reuse the tube for subsequent washes.
- Add 500 µL Wash Solution #2/3. Draw the wash solution through the filter as in the previous step.
- Repeat with a second 500 µL aliquot of Wash Solution #2/3.
- After discarding the wash solution, continue centrifugation, or leave on the vacuum manifold for ~10–30 seconds to remove the last traces of wash solution.
- Put the Filter Cartridge into a fresh Collection Tube.
- Pipet 40-60 µL of Elution Solution preheated to ~70–80°C to the center of the filter. Close the cap of the tube.
- Recover eluate by centrifugation for ~30seconds at room temperature (RCF 10,000–15,000 x g).
- Add a second aliquot of hot Elution Solution to the center of the filter and re-spin for ~30 seconds. Typically, this second elution is collected into the same tube as the first elution.

APPENDIX C. RNA Isolation from Fresh Tissue using QIAzol

Important points before starting

- Do not allow tissue to thaw during weighing or handling prior to disruption in QIAzol Lysis Reagent.
- Homogenized tissue lysates (in QIAzol Lysis Reagent, step 4) can also be stored at -70°C for at least 1 month
- To process frozen lysates, thaw samples at room temperature or at 37°C in a water bath until they are completely thawed and salts in the lysis buffer are dissolved. Avoid extended treatment at 37°C, which can cause chemical degradation of the RNA. After thawing continue with step 5.

1. Use 1 ml of QIAzol Lysis Reagent for up to 100 mg tissue. The volume of the tissue sample should not be more than 10% of the volume of QIAzol Lysis Reagent used. Calculate the correct amount and pipet it into a appropriate vessel from homogenization and subsequent centrifugation.

2. If the entire piece of tissue can be used for RNA isolation, place it directly into the QIAzol Lysis Reagent, and proceed immediately with step 3.
- If only a portion of the tissue is to be used, determine the weight of the piece to be used, and place it into the QIAzol Lysis Reagent and then proceed immediately with step 3.
- Frozen tissue should not be allowed to thaw during handling. The relevant procedures should be carried out as quickly as possible.

3. Homogenize immediately using a conventional rotor-stator homogenizer until the sample is uniformly homogeneous (usually 20-40 seconds)
- NOTE: Incomplete homogenization will lead to significantly reduced yields of RNA. Homogenization with roto-stator homogenizers generally results in higher total RNA yields than with other homogenization methods.
- For samples containing a relatively high content of fat, proteins, polysaccharides, or extracellular material, centrifuge the homogenate at 12,000 x g for 10 minutes at 4°C in order to remove insoluble material. Carefully transfer the supernatant into a new collection tube and proceed to step 3.

4. Place the homogenate on the benchtop at room temperature for 5 minutes. This step promotes dissociation of nucleoprotein complexes.

5. For every 1 ml of QIAzol Lysis Reagent used in step 1 add 0.2 ml of chloroform.
- Cap the homogenate securely and shake it vigorously for 15 seconds

6. Place the homogenate on the benchtop at room temperature for 2-3 minutes

7. Centrifuge at 12,000 x g for 15 minutes at 4°C.
- After centrifugation, the sample separates into 3 phases: an upper, colorless aqueous phase containing RNA; a white interphase; and a lower, red organic phase. The volume of the aqueous phase is approximately 60% of the QIAzol Lysis Reagent used in step 1.

8. Transfer the upper aqueous phase to a new collection tube. Fro every 1 ml of QIAzol Lysis Reagent used in step 1, add 0.5 ml of isopropanol. Mix thoroughly by vortexing.

9. Place the tube on the benchtop at room temperature for 10 minutes

10. Centrifuge at 12,000 x g for 10 minutes at 4°C

11. Carefully aspirate and discard the supernatant.
- The RNA pellet is often visible as a gel-like or white pellet at the bottom of the collection tube

12. For every 1 ml of QIAzol Lysis Reagent used in step 1, add at least 1 ml of 75% ethanol. Centrifuge at 7500 x g for 5 minutes at 4°C.
- If the RNA pellet floats or sticks to the side of the collection tube, bring it to the bottom of the tube by centrifuging at 12,000 x g for 5 minutes at 4°C.

13. Remove the supernatant completely and briefly air dry the RNA pellet. Do not dry the RNA using a vacuum.

Redissolve the RNA in an appropriate volume of RNAse free water. Purify the RNA following the RNeasy mini, midi or maxi RNA cleanup protocol.

APPENDIX D. RNA Isolation from FFPE – ABI Protocol RecoverAll Total Nucleic Acid Isolation

A. Before you begin

RNase precautions

Lab bench, pipettors, microtome blade, and cutting surface
Before working with RNA, it is always a good idea to clean the lab bench, pipettors, and sectioning equipment with an RNase decontami- nation solution (e.g., Ambion® RNaseZap® Solution, P/N AM9780).

Gloves and RNase-free technique
Wear laboratory gloves for this procedure; they protect you from the reagents, and they protect the RNA from nucleases that are present on skin. Use RNase-free pipette tips to handle the kit reagents, and avoid putting used tips into the reagent containers.

Prepare Wash Solutions
a. Add 42 mL of ACS grade 100% ethanol to the bottle labeled Wash 1 Concentrate. Mix well.
b. Add 48 mL of ACS grade 100% ethanol to the bottle labeled Wash 2/3 Concentrate. Mix well.
c. Cap the wash solution bottles tightly to prevent evaporation.
d. Mark the labels to indicate that the ethanol has been added.

The final solutions will be referred to as Wash 1 and Wash 2/3 in the procedure.

B. Deparaffinization

1. **Cut 5-20 µm FFPE sections to obtain the equivalent of ≤80 µm**
   - Use sections that were cut by a microtome from the interior of the paraffin block, to minimize nucleic acid damage by exposure to the atmosphere during storage
   - Place the equivalent of ≤80 µm of tissue slices (i.e., a maximum of four-20 µm, eight-10 µm, or sixteen-5 µm slices) in a 2 ml microcentrifuge tube.

2. **100% Xylene, mix, and incubate**
   - Xylene treatment completely removes paraffin from the sections
   - Add 1 ml of 100% xylene to the sample
   - Vortex briefly to mix
   - Heat the sample for 3 mins at 50°C to melt the paraffin

3. **Centrifuge**
    - Centrifuge the sample for 2 mins at room temperature and maximum speed to pellet the tissue
    - If the sample does not form a tight pellet, recentrifuge for an additional 2 mins. If a tight pellet still does not form, proceed with caution in next step.
    - Remove the xylene without disturbing the pellet. Discard the xylene. If the pellet is loose, you may need to leave some xylene in the tube to avoid removing any tissue pieces.

4. **Wash**
    - The ethanol washes remove xylene from the sample and accelerate drying of the tissue
    - Add 1ml of 100% ethanol (room temp) to the sample and vortex to mix
    - Centrifuge the sample for 2 mins at room temp and maximum speed to pellet tissue
    - Remove and discard the ethanol without disturbing the pellet
    - The ethanol will contain trace amounts of xylene, and must be discarded accordingly
    - Repeat the above steps to wash a second time with 1 ml of 100% ethanol
    - Briefly centrifuge again to collect any remaining drops of ethanol in the bottom of the tube. Remove as much residual ethanol as possible without disturbing the pellet.
    - Vacuum dry pellet for 10-15 mins

C. Protease Digestion

1. **Digestion for RNA-only isolation**
    - Add 400µl Digestion Buffer to each sample
    - Add 4 µl Protease to each sample
    - Swirl the tube gently to mix and to immerse the tissue. If tissue sticks to the sides of the tube, use a pipet tip to push it into the solution, or briefly centrifuge to bring the tissue down into the solution

2. **Incubation**
    - Incubate the sample in a heat block or a water bath for 3 hours at 50°C

D. Nucleic Acid Isolation

1. **Isolation Additive**
    - Add 480 µl Isolation Additive to each sample
    - Vortex the mix
    - The solution should appear white and cloudy after mixing

2. **Ethanol**
    - Add 1.1 ml 100% ethanol to each sample. Pipet in two aliquots of 550 µl to accommodate adjustable pipettors.
    - Mix each sample by pipetting up and down carefully. The solution should become clear at this point. Be careful when closing lids as the tubes will be near capacity.

3. **Filter**
    - For each sample, place a Filter Cartridge in one of the Collection tubes supplied.
    - Pipet 700 µl of the sample/ethanol mixture from step 2 onto the Filter Cartridge and close the lid
    - To prevent clogging of the filter, avoid pipetting large pieces of undigested tissue onto the Filter Cartridge.
    - Centrifuge at 10,000 x g for 30-60 sec to pass the mixture through the filter
    - Discard the flow-through, and re-insert the Filter Cartridge in the same collection tube
    - Repeat the above steps until all the sample mixture has passed through the filter (this should take 3 passes)

4. **Wash 1**
    - Add 700 µl of Wash 1 to the Filter Cartridge
    - Centrifuge for 30 sec at 10,000 x g to pass the mixture through the filter
    - Discard the flow-through and re-insert the Filter Cartridge in the same collection tube

5. **Wash 2/3**
    - Add 500 µl of Wash 2/3 to the Filter Cartridge
    - Centrifuge for 30 sec at 10,000 x g to pass the mixture through the filter
    - Discard the flow-through and re-insert the Filter Cartridge in the same collection tube
    - Spin the assembly for an additional 30 sec to remove residual fluid from the filter

E. Nuclease Digestion and Final Nucleic Acid Purification

1. **RNA Isolation**
    - Combine the following solutions (amount per reaction) to make the DNase mix (a master mix can be used if there is more than one sample):
        - 6 µl 10X DNase Buffer
        - 4 µl DNase
        - 50 µl Nuclease-free Water
    - Add 60 µl of the DNase mix to the *center* of each Filter Cartridge.

- Cap the tube and incubate for 30 mins at room temp (22-25 °C)

2. **Wash 1**
   - Add 700 µl of Wash 1 to the Filter Cartridge.
   - Incubate for 30-60 secs at room temperature.
   - Centrifuge for 30 secs at 10,000 x g.
   - Discard the flow-through, and re-insert the Filter Cartridge in the same Collection Tube.

3. **Wash 2/3**
   - Add 500 µl of Wash 2/3 to the Filter Cartridge.
   - Centrifuge for 30 secs at 10,000 x g
   - Discard the flow-through, and re-insert the Filter Cartridge in the same Collection Tube
   - Repeat these steps to wash a second time with 500 µl of Wash 2/3
   - Centrifuge the assembly for 1 min at 10,000 x g to remove residual fluid from the filter.

4. **Elute**
   - Transfer the Filter Cartridge to a fresh Collection Tube.
   - Apply 60 µl of Elution Solution or nuclease-free water to the *center* of the filter and close the cap
   - For RNA isolation, use room temperature eluent (22-25 °C)
   - Allow the sample to sit at room temperature for 1 min
   - Centrifuge for 1 min at maximum speed to pass the mixture through the filter. The elute contains the RNA
   - Store the nucleic acid at -20 °C or colder
   - (Optional) To store your sample for an extended period of time, or if a very small amount of nucleic acid was recovered, transfer the eluate to a non-stick tube to prevent loss of the nucleic acid.

APPENDIX E. **RNA Isolation from FFPE – Roche Protocol**
High Pure RNA Paraffin Kit

A. Before you begin

- Sample material: 5-10 µm sections from formalin-fixed paraffin embedded tissue
- During fixation in formalin intracellular RNAses become inactivated. However, RNA is degraded and cross-linked to proteins or inter- or intra-molecularly.
- Therefore FFPE tissue can be stored at room temp.
- It is recommended to use sterile disposable polypropylene tubes and tips in order to avoid RNase contamination.
- Wear gloves during the procedure.

B. Preparation of Working Solutions

- Proteinase K – Dissolve Proteinase K in 4.5 ml Elution Buffer, Store at –15 to -25°C
- Wash Buffer I – Add 60 ml absolute ethanol to Wash buffer, Store at +15 to +25°C
- Wash Buffer II – Add 200 ml absolute ethanol to Wash Buffer, Store at +15 to +25°C
- DNase I – Dissolve DNase I in 800 µl Elution Buffer and mix thoroughly, Store at –15 to -25°C

C. Deparaffinization

- To 5-10 µm sections in a 1.5 ml reaction tube add 800 µl Hemo-De, incubate 5 minutes and mix overhead during incubation several times.
- Add 400 µl absolute ethanol and mix. Centrifuge for 2 mins and maximum speed (12,000 – 14, 000 x g) and discard supernatant.
- Add 1 ml absolute ethanol and mix by overhead shaking. Centrifuge for 2 mins at maximum speed and discard supernatant.
- Blot the tube briefly onto a paper towel to get rid of ethanol residues. Dry tissue pellet for 10 mins at 55°C.

D. RNA Isolation

1) If necessary 3 preparations can be pooled after step 4. To one tissue pellet (deparaffinized as described above) add 100 µl Tissue Lysis Buffer, 16 µl 10% SDS and 40 µl Proteinase K working solution. Vortex briefly in several intervals and incubate overnight at 55°C.

2) Add 325 µl Binding Buffer and 325 µl absolute ethanol. Mix gently by pipetting up and down.

3) Combine High Pure filter tube and the collection tube and pipet the lysate into the upper reservoir.

4) Centrifuge for 30 s at 8000 x g in a microcentrifuge and discard the flow through. Steps 3-4 can be repeated, in case RNA needs to be pooled, with two more tissue pellet preparations.

5) Repeat the centrifugation at maximum speed in order to dry the filter fleece completely.
6) Add 500 µl Wash Buffer I working solution to the upper reservoir. Centrifuge for 15 s at 8000 x g, discard the flow-through.
7) Add 500 µl Wash Buffer II working solution to the upper reservoir. Centrifuge for 15 s at 8000 x g, discard the flow-through.
8) Add 300 µl Wash Buffer II working solution, centrifuge for 15 s at 8000 x g, discard the flow through.
9) Centrifuge the High Pure filter for 2 mins at maximum speed.
10) Place the High Pure filter tube into a fresh 1.5 ml reaction tube, add 90 µl Elution Buffer. Centrifuge for 1 min at 8000 x g.
11) Add 10 µl DNase Incubation Buffer, 10X and 1.0 µl DNase I working solution to the eluate and mix. Incubate for 45 mins at 37°C.
12) Add 20 µl Tissue Lysis Buffer, 18 µl 10% SDS and 40 µl Proteinase K working solution. Vortex briefly. Incubate for 1h at 55°C.
13) Add 325 µl Binding Buffer and 325 µl absolute ethanol. Mix and pipet into a fresh High Pure filter tube with collection tube.
14) Centrifuge for 30 s at 8000 x g in a microcentrifuge and discard the flow-through.
15) Repeat the centrifugation at maximum speed in order to dry the filter fleece completely.
16) Add 500 µl Wash Buffer I working solution to the upper reservoir. Centrifuge for 15 s at 8000 x g, discard the flow-through.
17) Add 500 µl Wash Buffer II working solution. Centrifuge for 15 s at 8000 x g, discrd the flow-through.
18) Add 300 µl Wash Buffer II working solution. Centrifuge for 15 s at 8000 x g, discrd the flow-through.
19) Centrifuge the High Pure filter for 2 mins at maximum speed.
20) Place the High Pure filter tube into a fresh 1.5 ml reaction tube. Add 50 µl Elution Buffer, incubate for 1 min at room temperature. Centrifuge for 1 min at 8000 x g to collect the eluated RNA.
21) The microcentrifuge tibe now contains the eluted RNA. Either use 10 µl of the eluted RNA directly in RT-PCR or store the eluted RNA at - 80°C for later analysis. Before photometrical determination of RNA concentration, centrifuge the eluate for 3 mins at maximum speed and transfer supernatant to a fresh 1.5 ml reaction tube without disturbing glass fibers at the bottom of the original tube.

APPENDIX F. **RT-PCR used for cell lines**

1. **Dosage of RNA:**
   - Refer to the quantity of RNA that was calculated by the Nanodrop. This dosage is given in ng/µl.
   - This dosage gives you the volume of RNA to use for doing a RT-PCR, meaning 2 µg and the volume of water that must be added to have a total volume of 11 µl.
   - For example, if you have 945.92 ng/µl of RNA in your isolation, for 2000ng you will have 2.11 µl of RNA. You must then add 8.89 µl of miliQ water to your sample to have a total of 11 µl.

2. **RT-PCR:**
   - Use PCR RNAse free tubes
   - X µl of RNA (see Dosage instructions above)
   - Y µl of miliQ water (as described above for total of 11 µl)
   - 1 µl of Random Hexamer
   - *Thus you have a total of 12 µl*

   - Place in PCR instrument with the following program:
     - 5 minutes at 70°C
     - Pause at 4°C

   - Then add 8 *µl* of the following mix to each tube:
     - 5x reaction buffer : 4 *µl*
     - Ribonuclease inhibitor : 1 *µl*
     - dNTPs 10mM : 2 *µl*
     - revertaid H minus : 1 *µl*

   - Place back in PCR instrument with the following program:
     - 1 hour at 42°C
     - 10 minutes at 70°C
     - Pause at 4°C

   You now have cDNA that can be used for Q-PCR

APPENDIX G. **RT-PCR used for Fresh and FFPE tissue**
High Capacity cDNA RT Kit (ABI)

1. Thaw the RNA and all kit components

2. Prepare the master mix of the reverse transcription according to the following table:

Components	2 µg	5 µg
10X Reverse Transcript Buffer	2 µl	5 µl
10X Random Primer or Gene Specific Primer***	2 µl	5 µl
25X dNTPs	0.8 µl	2 µl
Multiscribe Reverse Transcriptase Enzyme (50U/µl)	1 µl	2.5 µl
Nuclease free H_2O	4.2 µl	10.5 µl
RNA diluted in H_2O	10 µl	25 µl
TOTAL	20 µl	50 µl

*** Gene Specific Primer = 1 µl of each of the 25 genes + 1 µl of each of the 3 housekeeping genes pooled together and mixed using ABI Taqman Assays (have 3'+ 5' oligo + fluorescence). Thus, when using this protocol use 2 µl of this mix instead of the Random Hexamer

3. Place in PCR instrument with the following program
- 10 mins at 25°C
- 2 hours at 37°C
- Pause at 4°C

4. Conserve cDNA at 4°C if used within the next 24 hours or at -20°C if using a later date.

APPENDIX H. QPCR ABI Protocol

- In each well pipet:
  - 5 µl – 2X TaqMan FastMix
  - 0.5 µl – ABI TaqMan Assay (20X)
  - 1.5 µl – cDNA
  - 3 µl – H_2O
- Then the plate is run by the 7900HT RT-QPCR machine

APPENDIX I. Analysis of Results from Q-PCR

Definitions of the terms of the results

Ct = Cycle Threshold
Value at which point the PCR curve crosses the threshold. A qPCR has approximately 40 cycles. The higher the Ct (30-35), the less present is the detected mRNA, because it requires more PCR cycles to detect the fluorescent amplification. If the Ct is small (10-15), the gene is strongly expressed. The housekeeping genes often have a smaller Ct then the other genes.

Delta Ct = Ct gene – Ct housekeeping gene

Delta Delta Ct = Delta Ct sample1 – Delta Ct calibrator

Delta Ct SD = Standard Deviation
The Standard deviation is calculated by the software. This error reflects the quality of the triplicates technique for the gene being tested and the housekeeping gene for the same sample. For the value to be considered valid, the standard deviation should be under 0.25. If the standard deviation is higher than 0.25, the RQ value is considered not reliable.

RQ = Relative Quantification = $2^{-\Delta\Delta Ct}$
This is the fold change. The calibrator is fixed at a value of 1. The other samples have a value with regards to the calibrator. A value of 10 indicates that the expression of the gene is 10X more than the calibrator. A value of 0.1 indicates that the expression of the gene is 10X less than the calibrator. An RQ greater than 2 or less than 0.5 is considered significant.

APPENDIX J. **Q-PCR on LDA Protocol**

1. Prepare a master mix for each cDNA being tested

For Duplicates prepare the following according to this guide :

Components	48 genes on 96 well plate 20 predicitve genes (duplicates) 4 housekeeping genes (duplicates)
2X TaqMan mix	100 µl
cDNA (25ng/ µl)	10 µl
MilliQ H_2O	90 µl
Total	200 µl

2. Distribute (2X for 48 genes) 100 µl of the mix to each port site

3. Centrifuge 2 x at 1200 rpm for 1 minute

4. Seal the plate with the ABI plate sealer

5. Cut the excess seal around the plate with scissors

6. Run the plate on the 7900HT Real-Time qPCR or conserve at 4°C for a maximum of 48 hours

7. PCR program:

- 95°C for 10 minutes
- 95°C for 15 seconds (45 cycles)
- 60°C for 1 minute

APPENDIX K. **Consent Form and Research Protocol**

FORMULAIRE D'INFORMATION ET DE CONSENTEMENT ÉCLAIRÉ DESTINÉ À LA PATIENTE

Recherche sur échantillons biologiques et de données cliniques et biologiques sur le cancer du sein

Chercheurs Principaux :

Dre. Sylvie Mader, Ph.D. Institut de Recherche de l'Immunologie et Cancérologie de l'Université de Montréal.

Dr. André Robidoux, M.D. Chirurgien oncologue, Département de Chirurgie Oncologique, Hôtel Dieu de Montréal, CHUM

Préambule :

Vous êtes invitée à contribuer à cette étude parce que vous avez été opérée au sein dans le contexte d'un protocole de recherche du NSABP. Lors de votre consentement à ce protocole vous avez autorisé le laboratoire de pathologie de l'hôpital Hôtel-Dieu de Montréal du Centre Hospitalier de l'Université de Montréal à conserver vos échantillons biologiques de vos tissus mammaires en bloc de paraffine après votre chirurgie dans le but des recherches du NSABP. On vous demande si vous souhaitez participer à cette nouvelle recherche où on utilisera ces mêmes échantillons biologiques afin de répondre à nos nouvelles questions de recherche.

Également, les chercheurs peuvent avoir besoin de renseignements vous concernant. À cet effet, les chercheurs autorisés par ce projet de recherche consulteront votre dossier médical pour obtenir les informations indispensables pour leur permettre d'analyser les résultats de ces recherches. Elles incluent votre âge, votre sexe, votre groupe ethnique, votre diagnostic, votre état de santé actuel, l'existence d'antécédents médicaux personnels et familiaux, les traitements que vous avez reçus et comment vous y avez répondu. Ces informations seront informatisées en respectant les règles de confidentialité telles que prévue par la législation provinciale et fédérale.

1. DÉFINITIONS

En recherche biomédical, est appelé échantillon biologique tout organe, tissu ou autre substance biologique (par exemple le sang, urine, etc.) prélevé sur une personne. Le sérum représente la partie liquide du sang. Les lignées cellulaires sont des cellules provenant d'un échantillon biologique qui ont été traitées de façon à être capables de se multiplier en laboratoire.

Les gènes sont des molécules qui renferment toutes les informations génétiques transmissibles permettant un bon fonctionnement des cellules de notre corps, ainsi que les informations permettant à ces cellules de se défendre contre les agressions diverses dont notre corps fait l'objet. Nos gènes fournissent aussi l'information qui détermine nos caractères héréditaires, tels que la couleur de nos yeux ou notre groupe sanguin. Cette information génétique est portée par l'ADN. Il correspond à des phrases écrites avec un alphabet de 4 lettres, le code génétique, qui va être traduit en protéine par l'intermédiaire d'un messager (ARN). Ces protéines qui constituent notre organisme sont indispensables au bon fonctionnement de l'individu. Une anomalie, même d'une seule lettre dans la phrase, peut favoriser l'apparition de cancers. L'identification des ces anomalies permettra de mieux comprendre les mécanismes d'apparition de ces cancers.

2. JUSTIFICATION DU PROJET

Le cancer du sein est un problème de santé majeure pour les femmes. Dans 5-10% des cas, il peut exister une prédisposition familiale à développer un cancer. Des chercheurs travaillent dans ces domaines, à la recherche des causes de cette maladie qu'elle soit ou non d'origine familiale, ainsi qu'à la recherche de nouveaux traitements. Pour que ces recherches puissent se poursuivre, les chercheurs ont besoin d'échantillons biologiques provenant autant d'organes ou de tissus sains que d'organes ou de tissus présentant des problèmes bénins ou malins. De plus, les chercheurs peuvent avoir besoin que les cellules, obtenues à partir de votre échantillon biologique, soient mises en culture (formations de lignées cellulaires) de façon à disposer de suffisamment de matériel biologique pour réaliser les recherches. Dans certains cas, en particulier dans le cas de cellules normales ou proche de la normale, elles doivent être traitées pour être capable de se multiplier.

3. OBJECTIF DU PROJET

Les 2/3 des cancers du sein sont positifs pour le récepteur des œstrogènes (ER) et sont stimulés par les œstrogènes. Un traitement de Tamoxifen (anti-œstrogène) peut réduire le taux de rechute des cancers ER positive ainsi qu'augmenter la survie des femmes atteinte de ces cancers. Toutefois cette thérapie n'est pas efficace dans tous les cancers du sein ER-positives. L'hormonothérapie va bénéficier à environ 60-70% des cancers ER-positives. Nous avons

identifié 20 gènes qui peuvent prédire la réponse au Tamoxifen pour nous dire quelles femmes en bénéficient le plus. Nous voulons valider ces 20 gènes sur des tumeurs de cancer du sein traités ou non avec du Tamoxifen. Par la suite nous voulons développer un outil qui pourrait prédire la réponse au Tamoxifen. En faisant ceci on pourrait savoir à qui ce traitement est bénéfique et à qui le traitement est futile.

4. BÉNÉFICES

Il se peut que vous retiriez un bénéfice personnel de votre participation à ce projet de recherche, mais on ne peut vous l'assurer. Par ailleurs, les résultats obtenus contribueront à l'avancement des connaissances dans ce domaine.

5. RISQUES

Puisque vous étiez déjà opéré dans le passé et on a déjà vos tissus au laboratoire, il n'y a aucun risque supplémentaire pour votre santé si vous participez à cette étude. Dans le cas présent, nous allons prendre des mesures strictes en matière de confidentialité et de sécurité des données qui vont être contenues dans notre recherche ainsi qu'en matière de respect de la vie privée. Nous nous engageons à ne divulguer aucune information, ni à votre famille, ni à des assureurs ou employeurs, sans votre consentement écrit.

6. CONFIDENTIALITÉ

Durant votre participation à ce projet de recherche visant d'utiliser du matériel biologique, le chercheur responsable ainsi que son personnel recueilleront certains renseignements vous concernant de votre dossier médical.

Ces renseignements comprennent des informations concernant votre âge, votre sexe, votre origine ethnique, votre diagnostic, votre état de santé passé et présent, votre histoire familiale ainsi que des résultats de tous les tests, examens et procédures que vous auriez déjà passé.

Seuls les renseignements nécessaires pour répondre aux objectifs scientifiques de ce projet seront recueillis. Les chercheurs utiliseront les données et les échantillons biologiques dans le but de répondre à ces objectifs scientifiques.

Tous les renseignements et les échantillons biologiques recueillis au cours de ce projet demeureront strictement confidentiels dans les limites prévues par la loi. Afin de préserver votre identité ainsi que la confidentialité des renseignements et des échantillons, un numéro de code leur sera attribué. La clé du code reliant votre nom aux renseignements et aux

échantillons sera conservé par le Dr. André Robidoux. Votre nom et vos coordonnées ne feront pas partie des informations versées dans la banque de données et de matériels biologiques.

Vos échantillons de cancer du sein en bloc de paraffine sont conservée de façon indéfini tel que déjà prévu avec le NSABP, lors de votre consentement au protocole du NSABP.

Les données pourront être publiées dans des revues spécialisées ou faire l'objet de discussions scientifiques, mais il ne sera pas possible de vous identifier.

À des fins de surveillance et de contrôle, la banque de données et de matériels biologiques pourra être consultée par une personne mandatée par le Comité d'éthique de la recherche de l'établissement, par une personne mandatée par des organismes publics autorisés. Toutes ces personnes et ces organismes adhèrent à une politique de confidentialité.

Vous avez le droit de consulter la banque de données et de matériels biologiques pour vérifier les renseignements recueillis vous concernant, et les faire rectifier au besoin, et ce, aussi longtemps que le chercheur responsable du projet ou l'établissement détiennent ces informations.

7. COMMERCIALISATION

Votre échantillon biologique ne pourra être vendu. Par conséquent, il sera utilisé uniquement à des fins de recherche. L'analyse de celui-ci pourra contribuer à la création de produits commerciaux dont vous ne pourrez retirer aucun avantage financier. Ces avantages financiers seront partagés entre le chercheur et l'Université de Montréal selon un protocole d'entente convenu par les deux parties.

8. INFORMATIONS SUR LES RÉSULTATS

La recherche requiert beaucoup de temps. C'est pourquoi nous ne pourrons vous donner les résultats de ces recherches dans l'immédiat.

Toutefois, dans le cas où les travaux de recherche aboutiraient à des résultats scientifiquement validés, pertinents à votre condition et à un éventuel traitement, si vous le désirez, vous serez contactée par l'intermédiaire d'un professionnel de la santé. Si vous souhaitez connaître votre statut personnel regardant cette information, le résultat ne pourra vous être donné qu'après avoir été confirmé par un nouveau test biologique fait dans un cadre clinique.

Si l'information obtenue peut avoir un impact sur la santé d'autres membres de votre famille, et dans l'éventualité de votre décès ou d'une autre cause vous rendant incapable de vous

déplacer pour connaître ce résultat, vous pouvez autoriser un professionnel de la santé de contacter une ou plusieurs personnes de votre famille que vous désignerez.

9. LIBERTÉ DE PARTICIPATION ET DROIT DE RETRAIT

Votre contribution à cette recherche est tout à fait volontaire. Vous êtes donc libre d'accepter ou de refuser d'y contribuer sans subir de préjudice sur vos soins actuels ou futurs.

Vous pouvez changer d'avis à tout moment. Si vous ne souhaitez plus que votre échantillon biologique soit utilisé à des fins de recherche, vous n'avez qu'à contacter le responsable du projet de recherche.

Votre échantillon biologique sera alors garder comme tel dans la banque du NSABP et ne sera plus utilisé.

Cependant, tous les résultats qui auront été obtenus avec votre échantillon avant la date de désistement seront conservés. Par la suite, vous ne serez pas recontactée et ne pourrez pas être informée si la recherche aboutit à des résultats scientifiquement validés. Si vous vous retirez du projet, à ce moment là, votre dossier médical ne sera plus consulté à des fins de recherche.

10. PROBLÈMES OU QUESTIONS

Si vous avez des questions concernant la constitution de ce projet de recherche, veuillez contacter le responsable du projet, André Robidoux 514-890-8000, poste 15535.

Pour toute question concernant vos droits en tant que sujet participant à ce projet de recherche vous pouvez communiquer avec la commissaire adjointe à la qualité des services de l'Hôpital Hôtel-Dieu de Montréal : Michèle Morin 514-890-8000, poste 12761.

11. SURVEILLANCE ÉTHIQUE

Le comité d'éthique de la recherche du CHUM a approuvé ce projet de recherche et en assure le suivi. De plus, il approuvera au préalable toute révision et toute modification apportée au formulaire d'information et de consentement et au protocole de recherche.

12. CONSENTEMENT

J'ai lu le présent formulaire de consentement intitulé ''Recherche sur échantillons biologiques et de données cliniques et biologiques sur le cancer du sein''. J'ai eu l'occasion de poser toutes mes questions au sujet du projet de recherche, et les réponses obtenues ont été satisfaisantes.

Je comprends que ma participation au projet se fait sur une base entièrement volontaire, e que je demeure libre de me retirer du projet à tout moment et sans préjudice. Il est entendu que, quelle que soit ma décision, cela n'affectera en aucune façon les soins que je devrais recevoir. En signant le présent formulaire, je ne renonce à aucun des mes droits légaux ni ne libère les chercheurs, l'hôpital de leur responsabilité civile et professionnelle.

Je recevrai une copie signée et datée de se formulaire d'information et de consentement.

En conséquence, je consens aux points suivants :

A) Étude sur le cancer su sein :

- J'accepte qu'une partie de mon matériel biologique qui a été prélevé dans le passé pour le protocole du NSABP puisse être utilisé à des fins de recherches, et j'autorise l'accessibilité à mon dossier médical afin de fournir les informations nécessaires à l'utilisation de cet échantillon.

  ( ) OUI ( ) NON

B) Choix d'être recontactée :

- J'autorise le responsable du projet ou son délégué à me contacter par téléphone si des informations complémentaires étaient nécessaires.

  ( ) OUI Téléphone : _____ ( ) NON

- Je désire être recontactée si la recherche sur le cancer du sein aboutit à des résultats scientifiquement validés.

  ( ) OUI Téléphone : _____ ( ) NON

En faisant le choix d'être recontactée dans le cas où la recherche aboutirait à des résultats scientifiquement validés, je m'engage à tenir l'équipe de recherche informée de mes coordonnées.

En cas de non disponibilité, j'autorise le responsable du projet de recherche ou son délégué à contacter :

_____ lien de parenté :_____

No de téléphone : _____

Après avoir indiqué vos choix, veuillez inscrire votre nom et votre signature ci-dessous.

_____	_____	_____
Nom de la participante	Signature de la participante	Date
_____	_____	_____
Nom du témoin	Signature du témoin	Date

Je certifie qu'on a expliqué au signataire intéressé les termes du présent formulaire, qu'on a répondu aux questions qu'elle à posées à cet égard, qu'on lui a clairement indiqué qu'il reste à tout moment libre de mettre un terme à sa participation au projet de recherche et décrite ci-dessus. Une copie signée du présent formulaire de consentement lui a été remise.

_____	_____	_____
Nom du responsable du projet de recherche	Signature du responsable du projet de recherche	Date
_____	_____	_____
Nom de la personne qui obtient le consentement	Signature de la personne qui obtient le consentement	Date

APPENDIX L. **Approval of Scientific Committee**

CHUM

EXPÉDIÉ 25 NOV. 2009

COMITÉ D'ÉVALUATION SCIENTIFIQUE
Édifice Cooper
3981, boulevard St-Laurent, Mezz 2
Montréal (Québec) H2W 1Y5

Le 24 novembre 2009

Dre Sylvie Mader

a/s Mme Nicole Tremblay
Hôtel-Dieu du CHUM
Pavillon LeRoyer
Bureau 1-310

Objet : **09.179 – Approbation finale CES**

Développement d'un outil multigénique prédictif de la réponse à l'hormonothérapie en cancer du sein

Chère Docteure,

J'accuse réception, en date du 23 novembre 2009, de votre courriel datée du 20 novembre 2009 en vue de l'approbation finale de l'étude décrite en rubrique. À la lecture du document reçu, le tout est jugé satisfaisant.

Il est entendu que vous ne pouvez commencer le recrutement de sujets avant d'avoir obtenu l'approbation finale du comité d'éthique de la recherche.

Vous souhaitant la meilleure des chances dans la poursuite de vos travaux, je vous prie d'accepter, Docteur, mes salutations distinguées.

Gilles Soulez, M.D.
Président
Comité d'évaluation scientifique du CHUM

GS/lf

CENTRE HOSPITALIER DE L'UNIVERSITÉ DE MONTRÉAL

HÔTEL-DIEU (Siège social)
3840, rue Saint-Urbain
Montréal (Québec)
H2W 1T8

HÔPITAL NOTRE-DAME
1560, rue Sherbrooke Est
Montréal (Québec)
H2L 4M1

HÔPITAL SAINT-LUC
1058, rue Saint-Denis
Montréal (Québec)
H2X 3J4

APPENDIX M. **Approval Ethics Committee**

CHUM

COMITÉ D'ÉTHIQUE DE LA RECHERCHE
Édifice Cooper
3981, boulevard St-Laurent, Mezz 2
Montréal (Québec) H2W 1Y5

Le 7 décembre 2009

Docteur André Robidoux
Département d'oncologie

A/s Dre Sylvie Mader
Hôtel-Dieu du CHUM
LeRoyer – 1-310

Objet : **09.179 – Approbation finale CÉR**

Titre : Développement d'un outil multigénique prédictif de la réponse à l'hormonothérapie en cancer du sein.

Protocole : N/A

Chère Docteure,

J'accuse réception, en date du 7 décembre 2009 de votre message électronique ainsi que du formulaire d'information et de consentement modifié (version du 7 décembre 2009) en vue de l'approbation finale de l'étude décrite en rubrique.

À la lecture de tous les documents reçus, le tout est jugé satisfaisant. Je vous retourne sous pli une copie du formulaire portant l'estampille d'approbation du comité. Seul ce formulaire devra être utilisé pour signature par les sujets.

La présente constitue l'approbation finale, **valide pour un an à compter du 16 novembre 2009** date de l'approbation initiale. Je vous rappelle que toute modification au protocole et/ou au formulaire de consentement en cours d'étude, doit être soumise pour approbation du comité d'éthique.

Cette approbation suppose que vous vous engagiez :

1. à **respecter la présente décision**;
2. à respecter les moyens de **suivi continu** (cf guide du chercheur)
3. à **conserver les dossiers de recherche** pour une période d'au moins deux ans suivant la fin du projets afin de permettre leur éventuelle vérification par une instance déléguée par le comité;
4. à respecter les modalités arrêtées au regard du **mécanisme d'identification des sujets de recherche** dans l'établissement.

CHUM

Le comité suit les règles de constitution et de fonctionnement de l'Énoncé de Politique des trois Conseils et des Bonnes pratiques cliniques de la CIH.

Vous souhaitant la meilleure des chances dans la poursuite de vos travaux, je vous prie d'accepter, Docteure, mes salutations distinguées.

Isabelle Ducios, avocate /
Vice-présidente
Comité d'éthique de la recherche

ID/kb

p. j. : Formulaire de consentement, estampillé et approuvé

c.c. : Par télécopieur au Bureau des contrats/ Centre de recherche, Hôtel-Dieu du CHUM, Pavillon Masson– 514-412-7134

Voici les coordonnées de la personne-ressource pour ce projet :

Mme Karima Bekhiti

Téléphone : 514 890-8000, poste 14528
Télécopieur : 514 412-7394