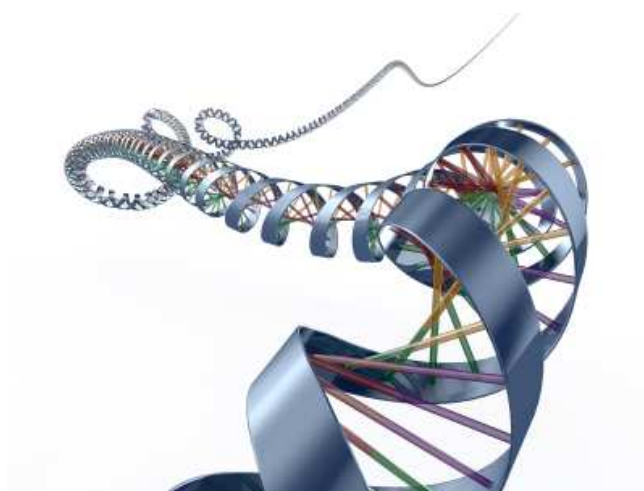


Aromatase and the Pharmacogenomic Profile

Influence of *CYP19A1* polymorphisms in the response of breast cancer patients treated with aromatase inhibitors



Heidi Schmid

Dissertação de Mestrado em Oncologia

2011

Heidi Schmid

Aromatase and the Pharmacogenomic Profile:

Influence of *CYP19A1* polymorphisms in the response of breast cancer patients treated with aromatase inhibitors

Dissertação de Candidatura ao grau de
Mestre em Oncologia submetida ao Instituto de
Ciências Biomédicas de Abel Salazar
Da Universidade do Porto

Orientador – Professor Doutor Rui Manuel de Medeiros Melo Silva
Categoria – Professor Associado com Agregação
Afiliação – Instituto Português de Oncologia
Francisco Gentil do Porto e Instituto de Ciências
Biomédicas Abel Salazar da Universidade do Porto

Agradecimentos

Antes de mais, gostaria de agradecer ao Prof. Dr. Rui Medeiros, coordenador do grupo de Oncologia Molecular do IPO-Porto, pela orientação, pelo rigor e disciplina e por toda a ajuda que me deu para conseguir realizar esta tese de Mestrado.

Ao Prof. Dr. Carlos Lopes e restante comissão coordenadora do Mestrado de Oncologia, assim como a todos os docentes deste mestrado, pelo empenho demonstrado e por tudo o que me ensinaram. À D. Maria do Céu por toda a disponibilidade e simpatia.

Ao Dr Laranja Pontes, Director do IPO-Porto, aos seus antecessores e colaboradores, pelo apoio concedido ao Mestrado de Oncologia e por me terem permitido realizar este trabalho no IPO-Porto.

À Liga Portuguesa Contra o Cancro – Centro Regional do Norte, bem como à Fundação AstraZeneca e ao Ministério da Saúde Português (CFICS-226/2001) pelo apoio e financiamento deste projecto.

Além disso, gostaria de deixar também um agradecimento muito especial ao Prof. Dr. Sobrinho Simões e ao IPATIMUP, pela colaboração neste projecto, especialmente à Dra. Ana Mafalda Rocha, pela competência, simpatia e rapidez demonstrada por ela.

Agradeço também à Dra. Idília Pina por me ter ajudado com a requisição e revisão dos processos clínicos.

Gostaria também de agradecer a todo o pessoal do Grupo de Oncologia Molecular, pelo apoio e por me animarem quando eu estava mais em baixo. Também ao pessoal do IBMC pelo apoio, amizade e tolerância.

Finalmente, mas não menos importante, um agradecimento muito especial ao meu namorado Nuno, por não me ter deixado desistir e por me ter apoiado em todo este difícil processo e por me ter feito acreditar que eu ia conseguir mesmo apesar de todas as dificuldades. Ao meu pai e irmão por me terem feito rir quando eu só queria chorar e por todo o carinho e amor que me deram ao longo deste trabalho. E um obrigado muito especial à minha Mãe, uma força da natureza, uma sobrevivente, que foi a minha fonte de inspiração para este trabalho.

True Perfection has to be imperfect...

(Oasis)

Abbreviation List

A

A: Adenine
AF-1: Activation Function 1
AJCC: American Joint Committee on Cancer
AKE: Hypotonic solution
AKT: Protein Kinase B (PKB)
ArKO: Knock-Out mice for Aromatase
ATAC: Arimidex, Tamoxifen, Alone or in Combination trial
ATM: Ataxia Telangiectasia Mutated

B

BIG 1-98: Breast International Group 1-98 trial
BRCA1: Breast Cancer 1, early onset
BRCA2: Breast Cancer 2 susceptibility protein

C

C: Cytosine
cAMP: Cyclic Adenosine Monophosphate
CBC: Complete Blood Count
COMT: Catechol-O-methyltransferase
COX-2: Cyclooxygenase 2
CYP: P450 Cytochrome
CYP19A1: P450 Cytochrome Family 19 subfamily A1

D

DHEAS: Dehydroepiandrosterone Sulphate
dNTP: Deoxiribonucleotide

E

EDTA: Ethylenediamine Tetraacetic Acid
EGFR: Epidermal Growth Factor Receptor
EORTC: European Organisation for Research and Treatment of Cancer trial
ER: Estrogen Receptor

F

FSH: Follicle Stimulating Hormone

G

G: Guanine

GST: Glutathione-S-Transferase

H

HDL: High Density Lipoprotein

HER2: Human Epidermal Growth Factor Receptor 2

I

IGFR: Insulin-like Growth Factor Receptor

IL: Interleukin

IL-6: Interleukin 6

IL-11: Interleukin 11

K

kb: kilobases

L

LDL: Low Density Lipoprotein

LH: Luteinizing Hormone

M

mRNA: messenger Ribonucleic Acid

N

NADPH: Nicotinamide Adenine Dinucleotide Phosphate

NAT: N-Acetyltransferase

NCBI: National Center for Biotechnology Information

O

ORF: Open Reading Frame

P

PBS: Phosphate Buffer Saline

PCR: Polymerase Chain Reaction

PGE2: Prostaglandin E2

PI3K: Phosphatidylinositol 3-Kinase

PKB: Protein Kinase B

R

Raf: Proto-oncogene Serine/Threonine-Protein Kinase

Rb: Retinoblastoma

RANK: Receptor Activator of Nuclear Factor Kappa-B

RANKL: Receptor Activator of Nuclear Factor Kappa-B Ligand

rpm: Rotations Per Minute

S

SERM: Selective Estrogen Receptor Modulator

SHBG: Sex Hormone-Binding Globulin

SNP: Single Nucleotide Polymorphism

T

T: Thymine

TARGET: Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability trial

TBE: Tris, Borate, EDTA

TNF: Tumour Necrosis Factor

TNM: Tumour size, lymph Node metastization and distant Metastases

TP53: Tumour Protein 53 gene

U

5'-UTR: 5' Untranslated Region

V

VNTR: Variable Number Tandem Repeat

INDEX

ABSTRACT	XI
RESUMO	XII
I. INTRODUCTION	13
1. BREAST CANCER	14
1.1. <i>Risk Factors for Breast Cancer</i>	14
1.2. <i>Estrogens</i>	17
2. THE AROMATASE ENZYME	21
2.1. <i>The CYP19A1 gene</i>	23
2.1.1. Expression of the CYP19A1 gene in different tissues	24
2.1.2. Regulation of the CYP19A1 gene in different tissues	26
2.2. <i>Importance of aromatase</i>	26
2.2.1. Aromatase Deficiency	28
2.2.2. Aromatase excess	29
2.2.3. Aromatase in breast cancer	30
3. AROMATASE INHIBITORS	32
3.1. <i>Third Generation Aromatase Inhibitors</i>	35
3.2. <i>Adverse Effects</i>	41
3.3. <i>Aromatase inhibitor resistance</i>	43
4. PHARMACOGENOMICS	45
4.1. <i>Polymorphisms</i>	45
4.2. <i>CYP19A1 polymorphisms</i>	46
II. AIMS	49
II. MATERIALS AND METHODS	51
1. POPULATION	52
2. SAMPLE PROCESSING	52
2.1. <i>Genomic DNA isolation</i>	53
3. GENOTYPING	53
3.1. <i>Polymorphism Selection</i>	53
3.2. <i>Genotyping</i>	53
3.3. <i>DNA Sequencing</i>	54
3.3.1. DNA Amplification	54
3.3.2. DNA Amplification Validation	55
3.3.3. Sequencing	55
4. STATISTICAL ANALYSIS OF THE RESULTS	56
III. RESULTS AND DISCUSSION	58

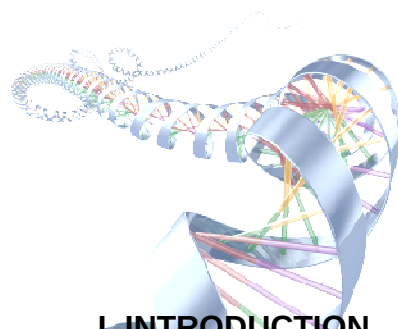
1.	STUDIED POLYMORPHISMS	60
1.1.	<i>Polymorphism Selection</i>	60
1.2.	<i>Allelic and Genotypic Frequencies Observed</i>	60
2.	POPULATION CHARACTERIZATION	64
2.1.	<i>Allelic distributions stratified for patient and tumor characteristics</i>	66
2.1.1.	SNPs	66
2.1.2.	Tetranucleotide repeat polymorphism	69
3.	SURVIVAL ANALYSIS.....	73
3.1.	<i>Global survival analysis</i>	73
3.2.	<i>Survival analysis for the SNPs</i>	76
3.2.1.	rs4646	76
3.2.2.	rs10046	80
3.2.3.	rs700518	83
3.2.4.	rs6493497.....	87
3.4.	<i>Tetranucleotide repetition polymorphism</i>	90
3.4.1.	(TTTA) _{7(-3bp)}	90
3.4.2.	(TTTA) ₁₁	94
3.4.3.	(TTTA) ₇	98
3.4.4.	(TTTA) ₈	102
4.	<i>Linkage Disequilibrium Analysis</i>	105
5.	<i>Overall Discussion</i>	110
V.	CONCLUSIONS AND FUTURE DIRECTIONS	113
VI.	BIBLIOGRAPHIC REFERENCES	116

ABSTRACT

The main aim of this study was to study the effects of the *CYP19A1* gene polymorphisms and the response to breast cancer treatment using aromatase inhibitors. In our study, 109 breast cancer patients were included, from whom we collected clinical data, in order to assess the associations between these data and the presence or absence of the alleles studied, as well as their effects on treatment response. The polymorphisms included in this study were rs4646, rs10046, rs700518 and rs6493497, as well as the tetranucleotide repetition polymorphism (TTTA)_n. We observed that all of the studied alleles had some effect on breast cancer response, although most of these effects were not conclusive. We observed the rs4646 and rs6493497 single nucleotide polymorphisms resulted in worse response to the treatment, conversely to rs10046 and rs700518, which improved the response. Concerning the tetranucleotide repetition polymorphism, we observed that the presence of eight and seven repetitions associated with a TCT deletion resulted in better response to the treatment, and that eleven and seven repetitions resulted in worse treatment response. Although all of these results need further confirmation, we observed a highly suggestive association between the eleven repetition allele and breast cancer treatment response. These results are very important in the clinical setting because they can have a strong impact on individualization of treatments, as well as in preventing side effects and absence of response.

RESUMO

O principal objectivo deste trabalho consistiu no estudo dos efeitos dos polimorfismos do gene *CYP19A1* na resposta ao tratamento do cancro da mama, utilizando inibidores de aromatase. Neste estudo foram incluídas 109 pacientes com cancro da mama invasor, de cujos processos clínicos foram colhidos dados em relação a características patofisiológicas e do tumor, de modo a analisar as possíveis associações entre os alelos estudados e essas mesmas características, bem como os seus efeitos na resposta ao tratamento. Os polimorfismos incluídos neste estudo foram rs4646, rs10046, rs700518 e rs6493497, bem como o polimorfismo de repetição de tetranucleótido (TTTA)_n. Os resultados obtidos sugerem que todos os alelos considerados têm algum efeito na resposta ao tratamento, apesar de a maioria desses efeitos não ter sido conclusiva. A presença dos alelos variantes de rs4646 e rs6493497 são sugestivamente associados a uma pior resposta ao tratamento, enquanto os alelos variantes de rs10046 e rs700518 são possivelmente associados a uma melhor resposta. No caso do polimorfismo de repetição de tetranucleótido, os resultados obtidos sugerem que a presença dos alelos de oito repetições e de sete repetições com uma deleção de TCT associada, resultam numa melhoria da resposta terapêutica. Por outro lado, a presença do alelo com onze repetições e do alelo com sete repetições originam piores respostas. Apesar de todos estes dados necessitarem de confirmação, foi observada uma associação altamente sugestiva entre o alelo contendo onze repetições de tetranucleótido e resposta ao tratamento com inibidores de aromatase. Estes resultados são muito importantes na clínica, uma vez que podem ter um grande impacto não só na individualização do tratamento, como também na prevenção de efeitos secundários e de ausência de eficácia.



I. INTRODUCTION

1. Breast Cancer

Breast cancer is the most frequent type of cancer among women in industrialized countries, accounting for about 20% of all new cases of cancer all over the world [1]. Besides, breast cancer is also one of the leading causes of death from cancer among women all over the world, and epidemiological studies suggest a slow but steady decrease in the age of occurrence [2]. Nowadays there are three times more new cases of breast cancer diagnosed annually compared with the 1960s, but the mortality remains very much unchanged. This increase in incidence but not in mortality can be due partly to the earlier detection of the disease and the availability of more efficient treatments [1].

1.1. Risk Factors for Breast Cancer

It is known that breast cancer is a very heterogeneous disease, and various risk factors have already been identified, such as age, hereditary, dietary (diet and obesity), gynecological (oral contraceptives, hormone replacing therapies, endogenous hormone levels, age of menarche and menopause, parity and mammographic density) and lifestyle (physical activity, smoking and alcohol) factors, among others, such as reactive oxygen species, radiation and environmental pollutants. Those risk factors will be described next.

Concerning age it is known that breast cancer incidence increases dramatically up to the age of 50, from which point this increase slows down [1].

Additionally, hereditary factors are observed in approximately one fourth of the total cases of breast cancer. It is believed that breast cancer initiation is a consequence of the accumulation of genetic damages which results in the activation of proto-oncogenes and inactivation of tumor suppressor genes. These genetic damages are followed by inappropriate cellular proliferation and/or aberrant apoptosis [1].

There are two classes of genes involved in hereditary predisposition: high penetrance genes, with allelic variants that are relatively rare, such as the breast cancer 1, early onset gene (*BRCA1*), breast cancer 2 susceptibility gene (*BRCA2*), tumor protein 53 gene (*TP53*) and *ataxia telangiectasia* mutated gene (*ATM*). However, these mutations only account for about 5% of the total breast cancer cases. On the other hand, low penetrance genes, such as the genes encoding for the enzymes involved in estrogen and carcinogen metabolism, as well as in the detoxification of reactive oxygen species

arising in these reactions, for instance P450 cytochromes (CYPs), Glutathione-S-transferases (GSTs), N-Acetyltransferases (NATs), and Cathecol-O-Methyltransferases (COMT), are more common and allelic variants confer low risk of breast cancer, contributing to sporadic cases, which result from the interactions with environmental factors [1]. Since high penetrance genes only account for a small proportion of breast cancer cases, relatively common genes acting together with endogenous risk factors, such as lifestyle, or low penetrance genes probably contribute to a larger proportion of breast cancer cases, as well as other high penetrance genes still unidentified [1].

Concerning dietary factors, the human diet contains a variety of natural carcinogens and anticarcinogens [1]. The increase in fat consumption, especially polyunsaturated fatty acids [3], as well as meat consumption [4], increase the risk of breast cancer. On the other hand, the uptake of fruits and vegetables, which are rich in antioxidants, reduces breast cancer risk [5].

Obesity has been associated to an increase in estrogen levels and breast cancer risk in postmenopausal women, who have most of their estrogens derived from the conversion of androgens, in the adipose tissue, as a result of the activity of the aromatase enzyme [6]. On the other hand, in premenopausal women, it is believed that obesity can have a protective effect, due to the higher period of anovulation frequently observed in these individuals, which results in a reduction of the estrogen levels [7].

A lifestyle factor associated breast cancer risk is physical activity, which is considered protective against breast cancer because it reduces regular ovulatory cycles and increases the levels of catechol-O-methylated estrogens [8]. In a case control study, the frequency of individuals who reported being more physically active was higher in healthy individuals than in carriers of breast cancer [9].

Concerning cigarette smoking, the results are controversial; being the risk of breast cancer only weakly associated to smoking [10]. However, some agents that constitute the cigarette smoke may have anti-estrogenic effects, such as nicotine, which has already been demonstrated to inhibit aromatase and also, women who smoke have an earlier menopause than nonsmokers, which can also be protective [11]. However, one has always to consider that the cigarette smoke is very rich in carcinogens and reactive oxygen species, which promote cancer [1].

It has already been demonstrated that alcohol increases the risk of breast cancer in a constant way, but only in about 15% of alcoholic women [12]. The exact mechanism by which alcohol induces breast cancer is not yet fully understood, but a number of studies suggest that women who drink alcoholic beverages have higher levels

of estrogens than women who do not drink [13]. Some proposed mechanisms consist of the production of reactive oxygen species induced by alcohol [14] and reduction of detoxifying protein expression [15].

The use of oral contraceptives increases breast cancer risk, which disappears after ten years of cessation [16]. On the other hand, concerning the use of hormone replacement therapy, it also increases breast cancer risk, but it disappears after five years of cessation, and breast cancer cases in these women tend to be less advanced at the time of diagnosis, and biologically less aggressive, compared to women who never used this type of therapy [17]. A case control study reports that breast cancer patients showed a higher probability of using hormone replacement therapy, compared to healthy individuals [18].

There is an association between endogenous estrogen levels and breast cancer risk [19]. The lifetime exposure to endogenous sex hormones is determined by some variables, which include age of menarche, age of the first full-term pregnancy, number of pregnancies and age of menopause [1]. Additionally, high estrogen levels in the serum or urine, and low levels of sex hormone binding protein (SHBG), resulting in high bioavailability of free estradiol also point for an important role for endogenous and exogenous estrogens in the risk of breast cancer [2]. The role of estrogens in breast cancer will be discussed later in more detail.

Additionally, the degree of mammographic density is a strong marker of breast cancer risk [20]. An epidemiological study has estimated that the risk in women with more dense breasts is four to six times higher than in women with less dense breasts [21]. However, the etiology of mammographic density is not completely understood, but evidence suggests that the exposure to steroid hormones can be an important factor, since it decreases with age, concomitantly with the period of menopause [22], as well as in women on tamoxifen [23], and also it increases in women who are on hormone replacement therapy [24]. One study based on the relationship of mammographic density with other breast cancer risk factors reports that it is negatively associated with body mass index, age, parity and age at menopause, and positively associated with age at menarche [20].

It has already been suggested that ionizing radiation increases the risk of breast cancer [25]. Additionally, an association was also found between the higher economic status and breast cancer risk since it affects parity, age at full-term pregnancy, diet and the use of synthetic hormones [26]. Besides, the history of benign breast disease is also a well-established risk factor for breast cancer [1].

On the other hand, the role of reactive oxygen species has also been associated to breast cancer etiology, since these compounds are mitogenic and have tumor promoting properties [1].

Environmental pollutants similar to hormones can also interfere in the control of a large family of nuclear hormone receptors, which in turn can up regulate various genes involved in the cell cycle, such as *TP53*, Retinoblastoma (*RB*) and the serine/threonine-protein kinase proto-oncogene *RAF* by transcriptional activation induced by ligand [2]. These environmental pollutants are designated xeno-estrogens, and include pesticides, dyes, food preservatives and other pollutants, and can play a role in the etiology of breast cancer, since they interfere with the activity of endogenous estrogens [27].

1.2. Estrogens

Although the etiology of breast cancer is very complex [28], most of the risk factors already discussed for breast cancer are associated with prolonged or increased exposure to estrogens, since it is considered that its main effect consists of the stimulation of breast cells, increasing the probability that a cell bearing a cancer causing mutation will spread [1]. Besides, estrogen exposure not only increases the risk, but is also associated to a worse prognosis of the disease [29].

The increased risk of breast cancer associated to early menarche probably results from a prolonged exposure of the breast epithelium to estrogens, early establishment of regular menstrual cycles and higher estrogen levels for more time and late menopause also increases the number of ovulatory cycles and consequently the exposure to estrogens. In fact for every year of menopause delay, the risk of breast cancer increases by 3% [1]. It has already been estimated that the duplication of estrogen levels in postmenopausal women increases the risk of breast cancer in approximately 30% [30]. These premises are supported by a case control study, in which patients had more frequently early menarche and late menopause than healthy individuals [9].

In addition, the higher parity and earlier age at first birth have also been associated to a decrease in breast cancer risk and this protective mechanism is not fully understood but probably reflects the earlier breast epithelial differentiation, rendering the cells less susceptible to suffer genetic damage. Also, prolonged lactation is protective in relation to breast cancer too [1]. In effect, a study reports that the frequency of women having the first child after 30 is significantly higher in breast cancer cases than in controls [31]. Another study also reports that the age of the first full-term pregnancy is a predictive

factor for breast cancer risk [32]. Yet another study also reports that the patients had a higher probability to be nulliparous or having children in later ages [18].

Steroids such as androgens and estrogens are essential components of the endocrine regulation of reproduction in vertebrates. Various diseases affecting fertility and general health are accompanied by aberrations in the metabolism of steroids, being frequently linked to infertility and reproductive failure, altered somatic growth and mineral metabolism, especially of calcium and in some cases cardiovascular problems. Therefore the balance between androgen and estrogen production is essential, not only for normal sexual development and reproduction, but also for normal growth and physiological well-being in both genders [33].

Estrogens are essential to women for reproductive organ development and in both genders for bone mineralization and gonad function [34]. Estrogens promote the growth and survival of normal and tumor breast epithelial cells, through the binding and activation of the estrogen receptor (ER). The activated ER, in turn, binds to nuclear gene promoters, activating many other genes involved in cell division, inhibition of apoptosis, angiogenesis and proteolytic activity. In fact, in the early stages of the breast precancer, an increase in ER expressing cells is often observed, as well as in 70% of all breast cancer cases [35]. Additionally, it has already been demonstrated that estradiol induces aneuploidy and structural chromosome changes, acting as a carcinogen [36].

In premenopausal women, almost all estrogens have an ovarian origin [35] acting on distant target tissues [37], while in postmenopausal women, most of the estrogens are derived from aromatization of androstenedione to estrone in the peripheral adipose tissue [1], as well as in estrogen sensitive tissues, such as breast, uterus, vagina, bones, brain, heart and blood vessels, as depicted in figure 1, in an autocrine fashion [35].

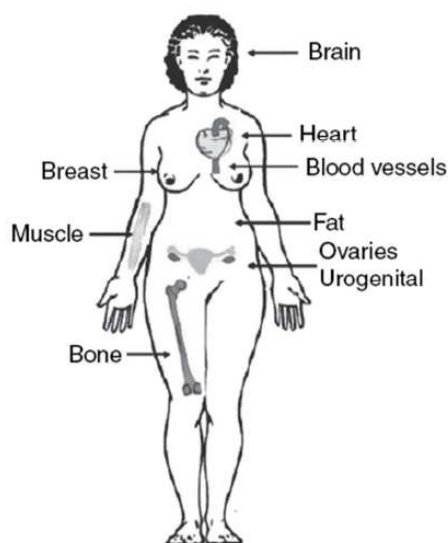


Figure 1: Main sites of aromatase activity and estrogen production [38].

Most of the estrogens produced locally in these extragonadal sites do not enter the circulation, but rather exert intracrine, autocrine and juxtacrine effects, acting directly on the cells in which they are produced or in neighboring cells. Thus, in the postmenopausal breast, the local estrogen biosynthesis determines its tissue levels, affecting the risk of breast cancer [38]. Furthermore, extragonadal sites of estrogen biosynthesis differ from the ovarian ones in the fact that estrogens produced in the first act predominantly on local tissues, and despite the levels of estrogens produced in these sites being lower than in the ovaries, the local tissue concentrations obtained are probably higher [39]. Therefore, estrogens can act directly and indirectly to induce breast cancer proliferation [40] and this is supported by a study that reports that 50 to 70% of estrogens result from local synthesis, with the rest being diffused from the plasma [41]. Additionally it has already been demonstrated that the estradiol levels in the breast tumor is at least twenty-fold higher than in the plasma [39] demonstrating the importance of estrogens in breast cancer carcinogenesis and progression [31].

There is a large number of studies suggesting that estradiol can increase the rate of development of new breast tumors [42], but the pathophysiological mechanisms underlying this process are not fully understood [9]. There are various theories that attempt to explain the mechanism by which estradiol promotes breast cancer: one of those theories, and the most accepted, suggests that estradiol stimulates cellular proliferation, with a concomitant increase in the number of cell divisions, and, while they divide, cells accumulate errors in DNA replication. Furthermore, as cells divide more quickly, the probability of occurrence of these errors increases and the time for repairing

them decreases. Thus estradiol initiates genetic mutations, which result in neoplastic transformation, causing the propagation of cells bearing these mutations, promoting also tumor growth. Another theory suggests that estradiol can be metabolized into 4-hydroxyestradiol and further to 3,4-estradiol quinone, which is a highly reactive metabolite with the ability of binding covalently to the DNA's adenine (A) and guanine (G), forming adducts, which are removed through the activation of glycosidase, resulting in a depurinated DNA segment, with weak links between the nitrogen bases, so the repair of this error-prone DNA can result in mutations which can initiate breast cancer. Finally, a third theory suggests a synergetic effect between receptor-mediated proliferation and adduct formation to originate and promote the breast tumor [42].

All estrogens have an aromatic A ring, a phenolic hydroxyl group in the third carbon and a methyl group in the thirteenth carbon. The main circulating estrogens are estradiol, which is the most active in breast tissues and has a hydroxyl group in the seventeenth carbon, and estrone, which has a ketone group in the same position. [1]

The estrogen biosynthesis involves a series of enzymatic reactions from cholesterol, to form androgens and estrogens, as shown in figure 2 [1]. Different forms of estrogen are synthesized from different androgenic substrates, in different tissues. For instance, estrone is produced from androstenedione, in the adipose tissue, while estradiol is produced from testosterone in the ovary granulosa cells, and estriol originates from 16 α -hydroxylated androgens in the placenta [43].

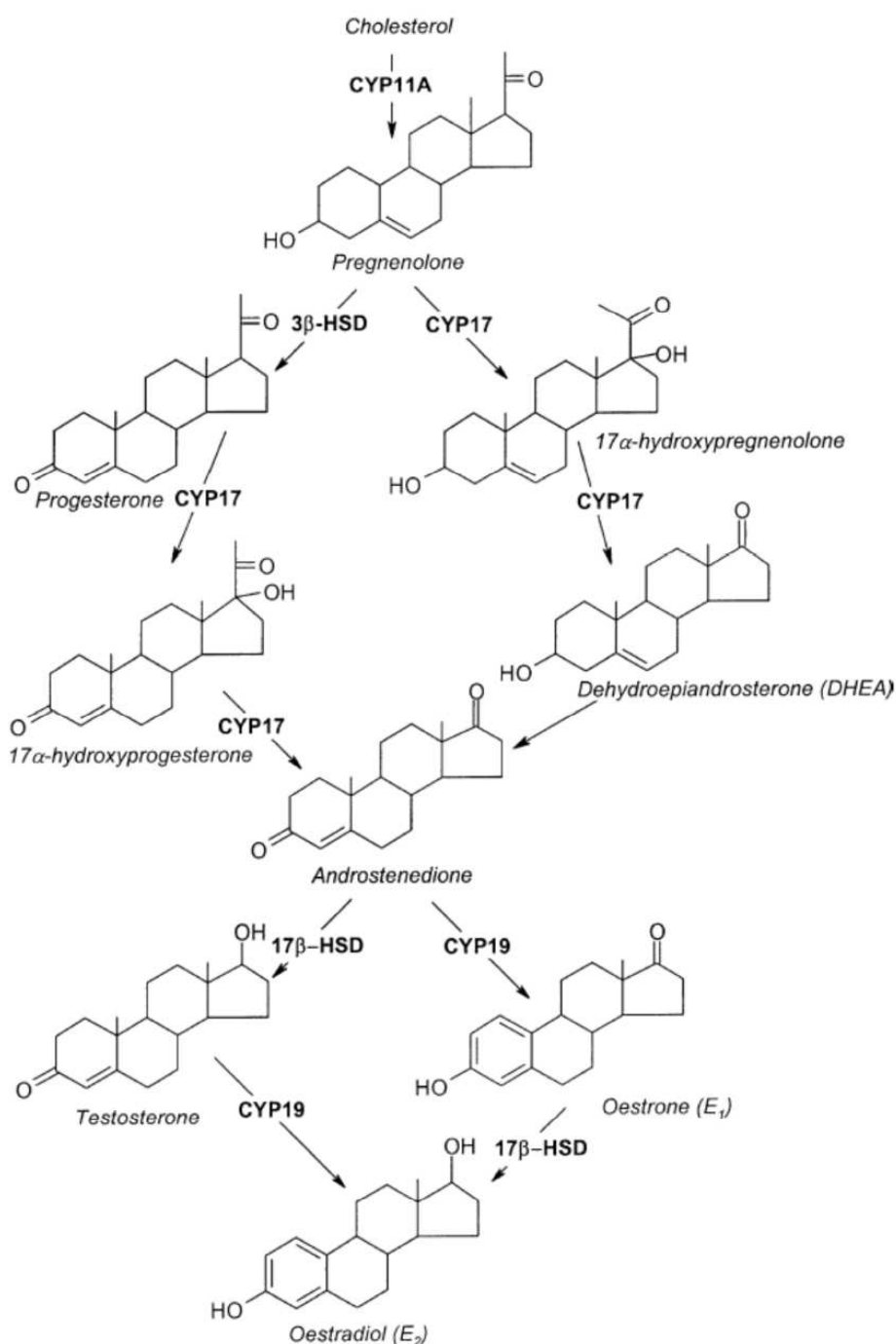


Figure 2: Schematics of estrogen production from cholesterol [2].

2. The aromatase enzyme

Aromatase belongs to the P450 cytochrome superfamily, which has this name derived from the wavelength of maximal absorption of ferrous carbon monoxide [38]. The members of this superfamily are involved in the detoxification of various xenobiotics and

in the phase I metabolism of a variety of drugs. Additionally, some isozymes belonging to this superfamily are also responsible for the activation of procarcinogens into carcinogens, especially in the polyaromatic hydrocarbon series [44]. This is a very large family of enzymes, containing more than 480 members, divided in 74 families, being aromatase the sole member of the 19 family [37]. The members of P450 cytochrome superfamily catalyze the incorporation of one oxygen atom into organic molecules, and for that reason are considered hydroxylases. Additionally, these enzymes are membrane proteins, and are present in most animal, plant and human tissues, such as the liver, and are essential for the synthesis of cholesterol and steroid hormones and for xenobiotic and fatty acid metabolism [38].

The common feature of the members of this family is the presence of a prosthetic heme group, which constitutes the active site. This heme group contains iron, which can be oxidized or reduced, with the help of the flavoprotein nicotinamide adenine dinucleotide phosphate (NADPH), which acts as a coenzyme for the reactions catalyzed by these enzymes [38].

The human aromatase enzyme is a 58 KiloDalton (kDa) protein, which is highly conserved in vertebrates. When bound to NADPH, forming a NADPH-P450 cytochrome reductase complex, it catalyzes a complex series of reactions that result in the conversion of androgens, with 19 carbon molecules, testosterone and androstenedione, to estrogens, with 18 carbon molecules, estradiol and estrone, respectively [38]. Hence, the estrogen biosynthesis is limited by the circulating androstenedione and testosterone levels [45].

Typically P450 cytochromes incorporate molecular oxygen in their substrates through reactions which are dependent on the efficient transfer of electrons from donor molecules. However, the catalytic process resulting in the aromatization of androgens to estrogens is particularly complex, as shown in figure 3 [33]. This reaction requires the sequential transfer of three pairs of electrons and consumes three moles of oxygen and three moles of reduced NADPH for the synthesis of one mole of estrogen [33]. It is believed that the reaction involves two consecutive hydroxylations on the methyl group present in the 19th carbon (19-hydroxylase activity) of the steroid substrate, originating 19-hydroxy- and eventually 19-oxo-intermediates. A third oxidative event, that is still debated, results in the cleavage of the methyl group of the 19th carbon (19-demethylase or 19-desmolase activity), aromatization of the steroid A ring and finally formation of estrogen [46]. The most common, recognized and physiologically important substrates are androstenedione and testosterone, but 16-OH-androstenedione, which is originated from the hepatic hydroxylation of the fetal adrenal dehydroepiandrosterone sulphate

(DHEAS), is an additional substrate used in the synthesis of estriol in the placenta of pregnant women and some primates [33].

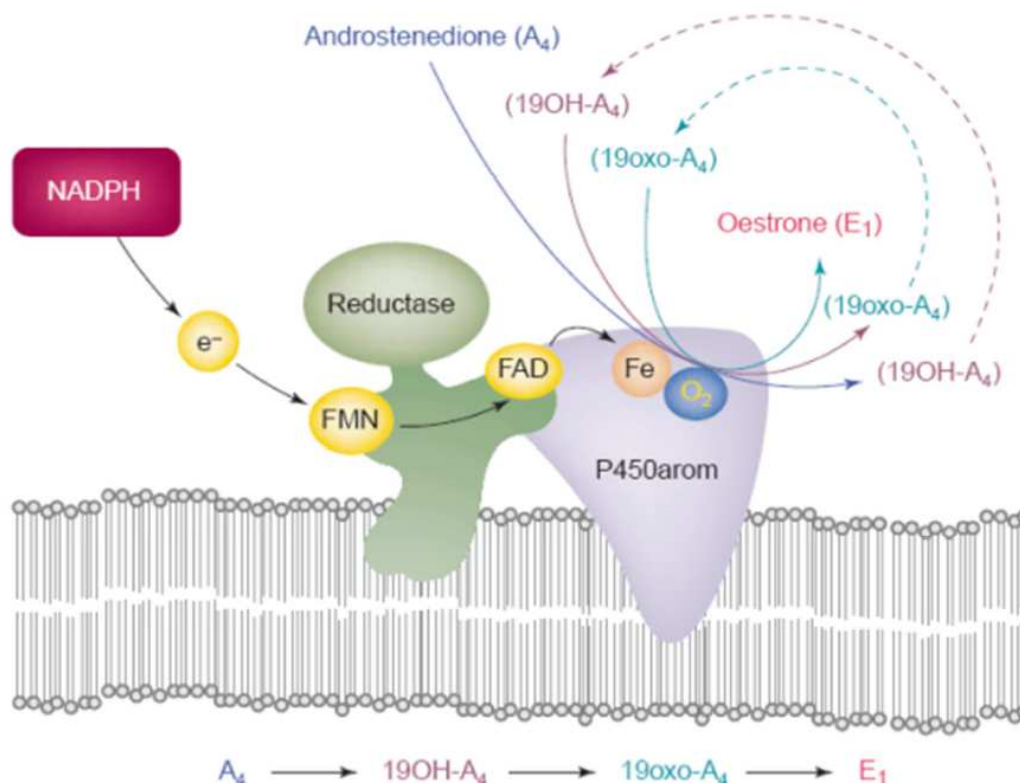


Figure 3: Structure of the NADPH-P450 cytochrome reductase complex and electron requirements and reaction [36].

Concerning its subcellular localization, it is believed that aromatase, as well as the reductase complex is present in the microsomal environment of the endoplasmic reticulum [47]. However, studies report also that the placental aromatase is associated to the mitochondrial compartment [33].

2.1. The CYP19A1 gene

Aromatase is encoded by the P450 Cytochrome family 19, subfamily A1 (*CYP19A1*) gene, localized in the short arm of the chromosome 15 [38], and was cloned for the first time in the 1990s [39].

The *CYP19A1* gene spans over about 123 kilobases (kb), with the coding region consisting of 30kb, while the 5' region spans over 93kb and functions as the regulatory unit of the gene [34]. This gene contains ten exons, from which nine are coding (exons II to X) [38], and the site of translation initiation (ATG codon) is localized in exon II, which is

centromeric. The 93kb of regulatory unit are telomeric, and contain various tissue-specific promoters [40]. The heme binding region of the aromatase enzyme is encoded by exon X [48].

2.1.1. Expression of the CYP19A1 gene in different tissues

In most mammals, aromatase is expressed in the gonads and brain, but the primates express the *CYP19A1* gene in additional extragonadal sites. The aromatase expression and estrogen production increase continuously as the phylogenetic tree evolves, reaching its highest in humans. This is due to the more efficient use of the existing promoters and recruitment of new tissue-specific promoters, in fat, skin, placenta and bone (Figure 1) [34]. Therefore, in humans, the expression of the *CYP19A1* gene is differentially regulated in different tissues by hormonally regulated promoters [2].

The levels of aromatase expression have interindividual and interregional differences and are also different in different stages of life. For instance, the fetal liver expresses aromatase, but the adult does not. Moreover, in women in reproductive age, the ovaries express high levels of aromatase, being these the main source of estrogen. On the other hand after the menopause the ovaries cease to function and peripheral tissues become important sites of estrogen production. Similarly, in men, 85% of estradiol and more than 95% of estrone are produced in extraglandular tissues, resulting from circulating androgen aromatization [38].

The key molecular mechanism which confers tissue-specific expression to the *CYP19A1* gene consists in the alternative use of various promoters, which regulate the mature messenger ribonucleic acid (mRNA) levels of aromatase, through alternative splicing of each exon I or 5'-untranslated region (5'-UTR), in a single splice junction immediately upstream of the coding region. Figure 4 represents the different tissue-specific exons I already identified [34].

I.4 (skin fibroblasts, preadipocytes and bone), I.5 (fetal), I.7 (endothelial and up-regulated in breast cancer) and I.f (brain), which locate between 33 and 95 kb from exon II [34].

However, since all mRNA species contain the same open reading frame (ORF), the encoded protein is always the same, independently of the promoter used [34]. Therefore, the use of alternative promoters does not affect the structure of the encoded protein, but rather its expression level [38].

2.1.2. Regulation of the CYP19A1 gene in different tissues

Since different tissues use their own promoters, enhancers and associated suppressors, the regulation of the *CYP19A1* gene in each tissue is extremely complex [38].

Some of the promoters do not contain the usual TATA and CAAT elements, and each one of those promoters is regulated in response to a series of specific hormones and cytokines [34]. For instance, in the gonads and brain, aromatase expression is regulated in different ways: while in the ovaries and testicles, the follicular stimulating hormone (FSH) and luteinizing hormone (LH) act through increasing concentrations of cyclic adenosine monophosphate (cAMP) to induce the expression of aromatase, in the brain the agents that cause an increase in the intracellular concentrations of cAMP, cause a decrease in aromatase activity, and androgens induce gene expression [33].

In the adipose tissue, on the other hand, expression of aromatase does not take place in adipocytes, but rather in the stromal cells that surround them, which can themselves be preadipocytes [49], through the adipose tissue promoter I.4, which regulates aromatase transcription in normal mesenchymal adipose cells at a relatively low level. This promoter is locally regulated by class I cytokines, such as interleukin 6 and 11 (IL-6 and IL-11, respectively) and oncostatin M, and tumor necrosis factor α (TNF- α) [38], and is stimulated by prostaglandin E2 (PGE 2) [37]. The main substrate of adipose tissue aromatase is the circulating androstenedione produced in the adrenal cortex [49].

2.2. Importance of aromatase

Initially, it was established that steroid hormones were only produced in endocrine glands, such as the ovaries, testicles and adrenal glands. However, in the 1970s it was demonstrated that the adipose tissue is a rich source of aromatase and that peripheral tissues are an important source of estrogens in men and postmenopausal

women. These studies determined that conversion of androstenedione to estrogen is higher in obese individuals, suggesting that adipose tissue might be an important site of aromatization [50]. Moreover, today it is also accepted that in most vertebrates, aromatase expression also takes place in the brain [49].

In many species, estrogen biosynthesis in the brain is implicated in sexual behavior [49]. A demonstration of that was the observation that sheep which prefer male partners showed increased aromatase activity in their sexually dimorphic nuclei [51]. A demonstration of the importance of aromatase expressed in the brain came from knockout mice for the aromatase gene that display a behavior that resembles obsessive-compulsive syndrome [52].

In the case of humans and some superior primates, there is a more extensive tissue distribution of estrogen biosynthesis, since it also occurs in the placenta and adipose tissue, as well as in bones, testicles and prostate [49].

In the end of the human pregnancy, 20mg a day of estradiol are produced by the placental aromatase. Estriol formation involves a complex collaboration between fetal adrenal glands and the placenta. The amount of aromatase in the liver increases substantially during pregnancy, probably due to the increase in cortisol levels. Together, these mechanisms result in a total estriol production of 100 to 150mg a day, which are 5 to 7 times higher than estradiol. Therefore, estriol might be the main estrogen during pregnancy, and is available to access the tissues because it does not bind strongly to sex hormone binding globulin (SHBG). Conversely, estradiol is less available, because it binds with high affinity to SHBG [48]. Although estriol is considered a weak estrogen, studies in sheep and mice uteri demonstrate their potent estrogen agonistic effects, as a stimulator of weight and blood flow in the uterus [53]. These factors are extremely important for optimal development and growth of the fetus, since they regulate the adequate supply of oxygen and nutrients through the hemochorial placenta to the fetal circulation during gestation. Additionally, the continued growth and development of the newborn depend on the increased blood flow to the breasts during pregnancy, in preparation for lactation, which is ensured by estrogens. All together, these data emphasize the key biological role of estriol, the main estrogen produced during pregnancy [48].

Alternatively, the bone aromatase is expressed mainly in osteoblasts and chondrocytes, being this local aromatase expression an important estrogen source, which is essential for bone mineralization [39]. Plasma estradiol levels correlate better with mineral bone density than testosterone levels in elderly men. This suggests an

important role for aromatized androgens in bone maintenance in men [54]. Besides, bone phenotype studies in men with aromatase and ER- α deficiency confirm this proposition, and the observed phenotypes include continued growth into adulthood, *genu valgum*, non-fused epiphyses and retarded bone age [55].

Aromatase is present in various compartments of the testicles of most mammals, such as Leydig cells, Sertoli cells, gonocytes, spermatogones, spermatocytes, spermatids and spermatozooids. However, the physiological significance of local estrogen production in the testicles is not yet fully understood [48]. Nevertheless, studies in knockout mice for *CYP19A1* demonstrate a defect in spermatogenesis, associated with a decrease in spermatozoid motility and inability to fertilize oocytes, suggesting a possibly important role for aromatase [56].

Concerning the prostate, studies suggest that aromatase locates predominantly in stromal cells in benign prostate hyperplasia and in both compartments in prostate cancer [57].

2.2.1. Aromatase Deficiency

Aromatase deficiency is a very rare recessive autosomal disorder, which results in various mutations in the coding region of the *CYP19A1* gene, causing reduced or even absent enzymatic function and, consequently, estrogen deficiency. Most of the already described mutations consist in single base changes in exons IX and X, which are important for substrate binding and encoding of the heme prosthetic group. These mutations originate changes in the codons and single aminoacid substitutions, or introduction of premature STOP codons, which originate truncated proteins [38]. The first cases of complete aromatase deficiency were identified in women with genital ambiguity and hypergonadotrophic hypogonadism in adolescence [33].

The studies in carriers of these mutations and the production of knock-out mice for the *CYP19A1* gene (*ArKO*), have allowed the detailed identification of some new unidentified estrogen functions, in males and females, not only related with sexual function [38].

In humans, in both genders, the first symptoms appear before birth, in the mother, which develops progressive virilization due to the inability to aromatize androgens in the placenta. The resulting excessive androgen levels in the uterus result in androgenization of the female fetus, which presents with ambiguous genitalia at birth. During childhood hemorrhagic cysts in the ovaries can occur and afterwards, during puberty, the adrenarche is normal, but primary amenorrhea occurs as well as absence of

mammary development. Moreover, as a consequence of androgen excess, acne and hirsutism can also take place, and virilization can progress with age [38].

Conversely, in men, the symptoms occur only after puberty, being the most distinctive one the linear progressive growth into adulthood, caused by incapacity of growth plate fusion, in the absence of estrogens, independently of the relatively high levels of testosterone. Besides, *genu valgum* is also present, eunuchoid body proportions, as well as retarded bone age, osteopenia and osteoporosis. This bone phenotype confirms the important role of estrogens in the maintenance of bone mass and maturation in men [38].

Another observed symptom in these patients is obesity and metabolic syndrome, defined by abdominal obesity, dyslipidemia – abnormal lipid levels in the blood – hyperinsulinemia and *acanthosis nigricans* which consists of skin darkening resulting from insulin resistance. Also, glucose intolerance or diabetes can occur and progress with age. Additionally, hepatic steatosis is also observed in some individuals, as well as abnormal fertility and loss of libido. Cryptorchidism, that is, absence of one or both testicles inside the scrotum has also been reported [38].

Concerning hormonal analyses, undetectable estradiol and estrone levels are observed, as well as increased levels of gonadotropins, while androstenedione and testosterone can be normal or increased [38].

2.2.2. Aromatase excess

There are few families described in the literature with estrogen excess syndrome, due to aromatase overexpression. The main features of this syndrome include severe pubertal gynecomastia in males and macromastia, premature puberty, enlarged uterus and menstrual irregularities in females. Moreover, premature growth plate fusion and low final stature are observed in both sexes [38].

Excessive androgen aromatization results in increased serum estradiol and estrone levels, with reduced levels of testosterone and androstenedione, as well as suppression of gonadotropin secretion [38].

Additionally, an increase in aromatase activity with age, obesity, hyperthyroidism, idiopathic gynecomastia and tumors, including testicular, adrenocortical, fibrolamellar hepatocellular and lung giant cell cancers, as well as melanomas, is also observed [38]. Therefore, aromatase excess is important in the etiology of various pathologies, and in all of them the proximal promoter used in the expression of aromatase is the same as the one controlling gonadal expression [33].

2.2.3. Aromatase in breast cancer

In estrogen-dependent pathological tissues, such as breast cancer and endometriosis, aromatase is overexpressed due to the inappropriate activation of aberrant promoters [34].

Breast cancer is highly dependent on estrogens for its development as demonstrated by the observation of high concentrations of ER in the breast tissue. [58] In fact, about 80% of breast tumors express ER at clinically significant levels for them to be considered ER-positive [59]. Additionally, CYP19A1 expression is higher in the breast tissue than in normal and surrounding breast tissues [60].

The breast adipose tissue is composed of mature cells containing lipids and other stromal components, being 90% of these cells fibroblasts – potential precursors of mature adipocytes – and 7% represented by endothelial cells [61]. Besides, studies suggest that aromatase activity resides mainly in adipose tissue fibroblasts [62].

Breast tumors also produce high levels of estrogen locally, through aromatase over-expression, and paracrine interactions between malignant epithelial cells, proximal adipose fibroblasts and vascular endothelial cells are responsible for estrogen production and absence of adipogenic differentiation in breast cancer tissues. Additionally, apparently, malignant breast epithelial cells secrete factors which inhibit adjacent adipose fibroblast differentiation into mature fibroblasts and also stimulate aromatase expression in undifferentiated adipose fibroblasts [63].

This adipose lack of differentiation in fibroblasts is in the origin of the desmoplastic reaction, that is, formation of a dense adipose fibroblast layer around malignant epithelial cells, which is essential for the structural and biochemical support during tumor growth. In fact, pathologists defend that approximately 70% of breast carcinomas are scirrhous, or have a hard consistency, like stone. This consistency is due to undifferentiated adipose fibroblast packaging surrounding malignant epithelial cells, which results in secretion of high quantities of TNF- α and IL-11, that inhibit adipose fibroblast differentiation in mature fibroblasts. Therefore, large quantities of these estrogen producing cells are maintained in the proximity of the tumor. Simultaneously, a series of different factors secreted by malignant epithelial cells activate aromatase expression in adjacent adipose fibroblasts, through activation of aberrant promoters in the tumor tissue and also in adjacent adipose fibroblasts [40].

The adipose tissue adjacent to the breast tumor, including the thick fibroblast layer seems to contribute to the majority of aromatase expression in breast cancer for various reasons. First, the amount of adipose tissue that surrounds a clinically detectable

tumor is comparatively very high. Furthermore, stronger positive aromatase immunostaining is observed in fibroblasts as compared to adipose tissue inside and surrounding the tumor fibrous capsule, as a result from the desmoplastic reaction [64]. Finally, aromatase expression levels and activity in fibroblasts isolated from breast adipose tissue or tumor are 10 to 15 times higher than those found in malignant epithelial cells [65], as shown in figure 5.

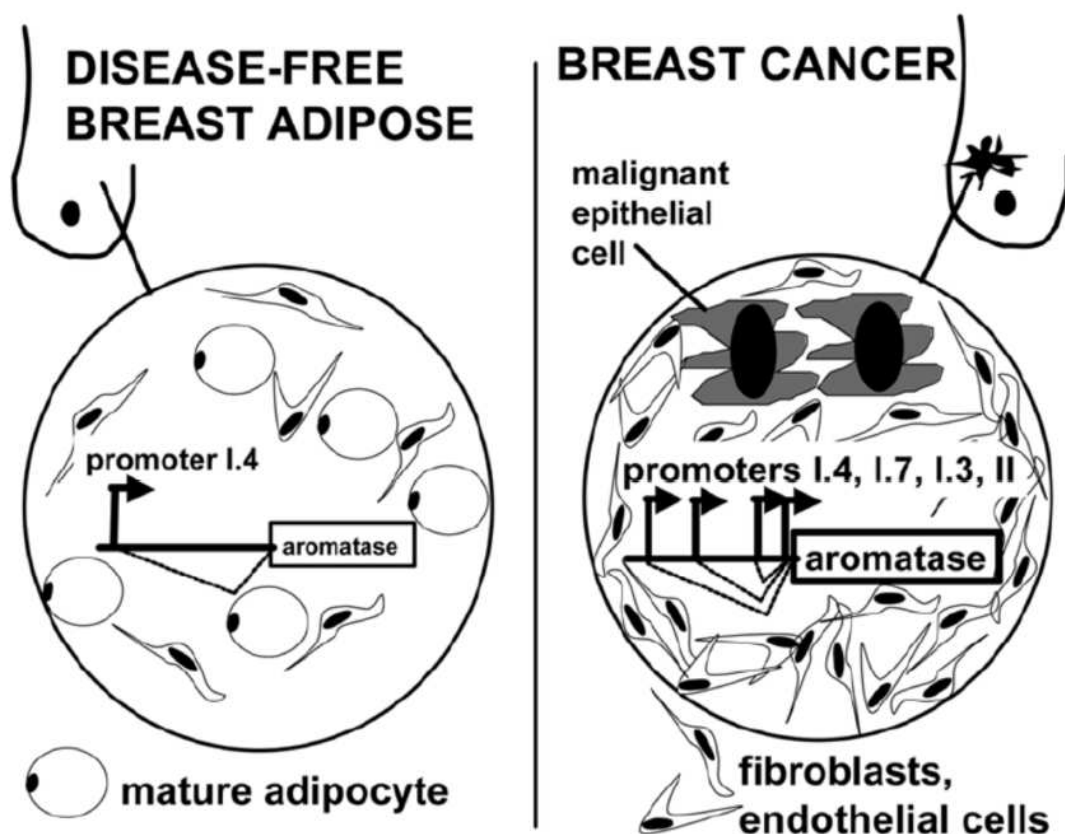


Figure 5: Promoter usage and distribution of adipocytes, fibroblasts and endothelial cells in disease-free breast adipose tissue (left) and breast cancer (right) [37].

Class I cytokines, such as IL-6, IL-11 and oncostatin M, and TNF- α , produced locally in adipocytes are important factors regulating promoter I.4, which, as already stated, regulates the expression of aromatase in normal adipose mesenchymal cells at a relatively low level. On the other hand, promoter PII is regulated by cAMP and gonadotropins. However, in the presence of breast cancer, which secretes various regulatory factors, stromal adipose cells can begin to use predominantly the PII promoter, together with promoters I.3 and I.7. This switch in promoter usage depends on the tissue microenvironment and results in an increase in aromatase transcription, expression of the protein and enzymatic activity. This process is the basis for the

increased estrogen production in stromal adipose cells which surround the breast tumor [38]. Evidence suggests that this switch from promoter I.4, adipose, to PII, ovarian, is due to alternative splicing in breast cancer patients with metastatic lymph nodes [66]. An alternative mechanism also proposed consists in the transcriptional activation of this promoter through induction or repression of regulatory elements upstream to the promoter, normally functional in all gonadal tissues [2].

Moreover, many breast tumors have the ability to over-express cyclooxygenase-2 (COX-2), producing prostaglandin E2 (PGE2), a powerful stimulator of aromatase expression, acting through cAMP, which regulates promoter PII regulated transcription in surrounding preadipocytes [38].

3. Aromatase Inhibitors

As previously acknowledged, estrogens promote normal and tumor breast epithelial cells as a result of binding to ER and its activation. The activated ER, in turn, binds to nuclear gene promoters, thus activating many other genes involved in cell division, apoptosis inhibition, angiogenesis and proteolytic activity. In early stages of breast pre-cancer, an increase in ER expressing cells is observed, as well as in 70% of breast cancer cases [35].

There are three strategies that can be used nowadays to impede this process (Figure 6). One of those consists in interfering with estrogen binding to the ER and/or promoter elements of genes regulated by it using selective ER modulators (SERMs) or anti-estrogens, such as Tamoxifen (Solfadox®) or Raloxifene (Evista®) (Figure 6 – A). Another strategy is the reduction or elimination of ER expression, with Fulvestrant (Faslodex®), which reduces the amount of ER available for estrogen binding (Figure 6 – B). Finally, the amount of estrogens can also be reduced by interfering with their production, though ovarian ablation in premenopausal women or aromatase inhibitors in postmenopausal women [35].

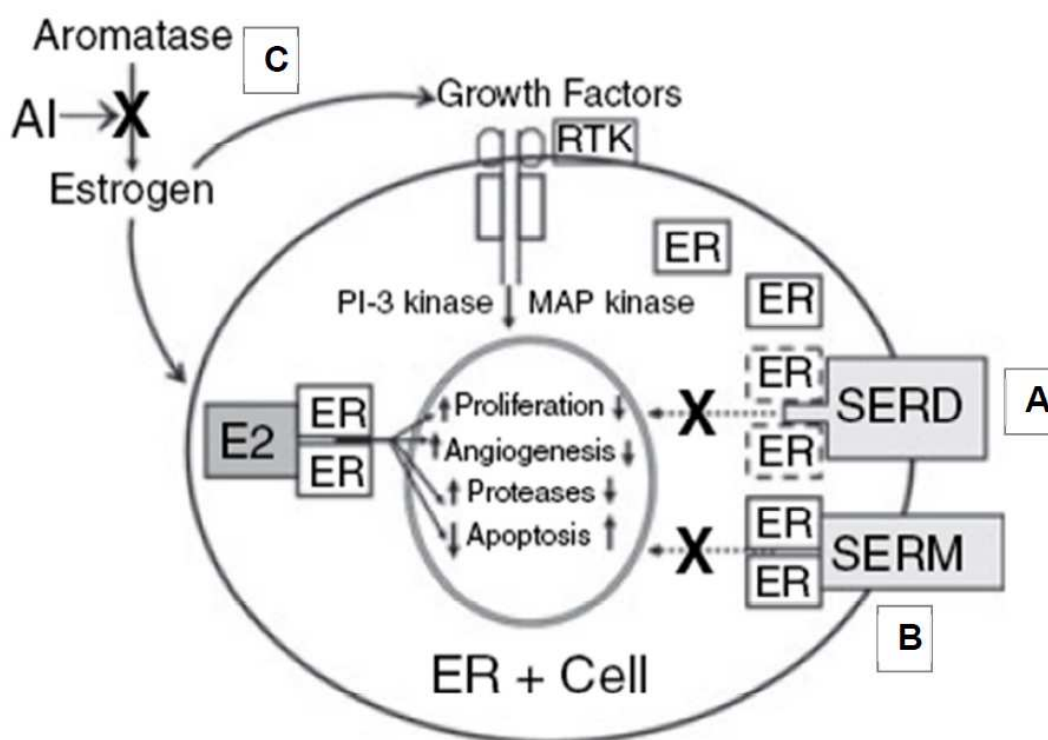


Figure 6: Effects of ER activation and strategies to interfere with this process. (A) Blocking of ER; (B) Reduction/elimination of ER expression; (C) Interference with estrogen production. (Adapted from [38])

In 1896, bilateral oophorectomy was the first established method as an effective treatment for advanced premenopausal breast cancer. Since then various other strategies were introduced with the intention of estrogen privation, such as hypophysectomy, surgical and later pharmacological adrenalectomy, selective blockade of ER (SERMs) and finally the use of potent aromatase inhibitors [38].

The fundamental choice for most of the patients with advanced breast cancer is between cytotoxic chemotherapy and hormonal therapy, being the latter preferred in the case of hormone receptor positivity. However, if the disease is rapidly progressing, chemotherapy might be the most adequate choice for initial treatment, since it acts more rapidly than hormonal therapy. However, hormonal therapy has the advantage of offering antitumor activity without the adverse effects inherent to chemotherapy, which result in a substantially worse quality of life and this issue is especially relevant for patients in palliative settings. Besides, the availability of various endocrine agents with different mechanisms of action means that cross resistance between agents is not a significant problem, and also allows the use of these different agents sequentially, permitting a

prolongation of the time in which agents can be used, and delay the need for cytotoxic chemotherapy [67].

Since its introduction in 1973 [68], tamoxifen has been the gold standard hormone therapy for treatment of breast cancer positive for hormone receptors, in patients who have never been exposed to this drug before or have a long interval between the last administration and metastatic presentation [69].

Tamoxifen is a selective estrogen receptor modulator, which competitively inhibits the binding of estradiol to ER, hindering in this way a series of mechanisms regulating cellular replication. The inhibition caused by tamoxifen alters the growth factor profile in responding tissues, causing arrest in the cell cycle, and changing the tumor proliferation and apoptosis processes, whose balance reflects on the antitumor responses and survival from the disease. Therefore, the ER and progesterone receptor levels correlate with global tamoxifen response [67]. Moreover, it has already been demonstrated that tamoxifen results in higher tolerability than hypophysectomy [70].

Treatment with tamoxifen during five years results in a reduction of the disease recurrence and mortality, being this benefit restricted to patients with ER-positive breast cancer. Also, this benefit persists for ten years of follow-up. Additionally, the advantage of tamoxifen is independent from lymph node involvement, age, dose, menopausal status or concomitant use of chemotherapy [71].

Nonetheless, although tamoxifen is a good breast cancer growth inhibitor, its effect on the human body varies and has mixed estrogenic properties. Most of these properties are desirable, including bone mineralization preservation in postmenopausal women and reduction of low density lipoproteins. However, these effects can be harmful and increase the risk of thromboembolic events and endometrial cancer [72]. Therefore, the prolonged exposure to tamoxifen confers a worse prognosis than to discontinue the treatment after five years [71]. This limitation can in part be based on the partial estrogen agonistic effect of tamoxifen, which contributes directly to the proliferation of endometrial cells, increasing the risk of endometrial cancer, and breast tumor stimulation [68]. Therefore, about a third of ER-positive breast cancers recur in spite of the adjuvant treatment with tamoxifen, with or without concomitant chemotherapy [35]. The reason for this agonistic effect consists in the fact that breast cancer cells are able to adapt in response to the increase pressure exerted by the treatment. The C-terminal region of α and β ERs contains a region of activation function 1 (AF-1), which can result in agonistic effects on ER-mediated transcription, due to loss in the balance between coactivators, corepressors and receptor integration and binding proteins. In this way, in some patients, these factors can allow that tamoxifen exerts estrogen agonistic effects [42].

The initial finding of aromatase and its various functions represents a great success in Endocrinology because it is the first molecular target for the rational development of drugs in breast cancer treatment [48]. The first aromatase inhibitors emerged in the 1960s [42], and acted directly in adrenal steroid synthesis [48]. Since then, three generations have already been developed. The first aromatase inhibitor that emerged was aminoglutethimide (*Cytadren*®) which was first used for pharmacological adrenalectomy, and later was found that blocked total body aromatization and was defined as a first generation aromatase inhibitor [38]. Aminoglutethimide was used for the treatment of hormone receptor positive breast cancer, with similar efficiency to tamoxifen, but with grave adverse effects and toxicity that reduced its utility [42].

Later emerged 4-hydroxiandrostenedione, an androstenedione analogue [38] or formestane (*Lentaron*®), a second generation steroid aromatase inhibitor, in the early 1990s. However it required intramuscular administration every two weeks and caused reactions in the injection site [68]. Therefore, formestane was used as a second line treatment after tamoxifen and was proven to be comparably effective and its adverse effects less serious than those of the first generation aromatase inhibitors. Another second generation aromatase inhibitor is fadrozole (*Afema*®). However, its use was restricted because of its rapid clearance and inhibition of aldosterone synthesis [38].

The two first aromatase inhibitor generations were limited by their side effects, which included hot flashes, dizziness, ataxia, nausea and vomits [35]. These limitations stimulated the development of third generation aromatase inhibitors [68], which will be discussed next.

3.1. Third Generation Aromatase Inhibitors

The lack of efficiency of first and second generation aromatase inhibitors and the limitations of tamoxifen encouraged the development of third generation aromatase inhibitors, which are compounds 100 to 1000 times more potent than the earlier compounds [42].

Aromatase inhibitors consist of a group of drugs with the ability of interrupting estrogen production through the inhibition of its conversion from androgens [38], reducing the amount of ligand available for ER [72]. This inhibition of androgen conversion into estrogens is achieved through aromatase enzyme suppression [35], preventing estrogen stimulation from reaching the tumor [73]. In addition to the suppression of circulating estrogen levels, aromatase inhibitors also abolish the autocrine

and paracrine estrogen production from peritumoral stromal cells in the primary site, as well as in metastatic disease [71].

Third generation aromatase inhibitors are superior to the previous generations because they are associated with less adverse effects and a more effective enzyme suppression [35], as well as higher specificity for the enzyme and time of action sufficiently long for them to be administered once a day [42]. Besides, as opposed to second generation aromatase inhibitors, such as formestane, they are highly specific, and have almost no effect on aldosterone and cortisol [71]. The results from studies comparing each third generation aromatase inhibitor, with aminoglutethimide/hydrocorticoids and progestin megestrol acetate, were almost consensual, showing the superiority of aromatase inhibitor efficacy [48]. Table 1 compares the three generations of aromatase inhibitors in terms of dose requirements and percentage of inhibition.

Table 1: Comparison of the three aromatase inhibitor generations in terms of dose requirements and percentage of inhibition. (Adapted from [40])

Drug	Dose	% of inhibition
1st generation		
Aminoglutethimide	1g	91
2nd generation		
Fadrozole	2 mg	82
Vorozole	1 mg	93
3rd Generation		
Letrozole	2,5 mg	99
Anastrozole	1 mg	97
Exemestane	25 mg	98

Table 1 shows that the mean inhibition degree inherent to the three third generation aromatase inhibitors is higher than 97%, with a low dose requirement, as compared to the two previous generations [38].

Moreover, there are also evidence that aromatase inhibitors are more efficient than tamoxifen, in ER-positive, progesterone receptor (PgR)-positive and/or with human epidermal growth factor receptor 2 (*HER2*) amplification tumors, as stated earlier [74]. As opposed to tamoxifen, aromatase inhibitors act through suppression of the conversion of androgenic substrates into estrogens, reducing the incidence of the serious secondary effects associated to tamoxifen, and maintaining the efficacy of the treatment on hormone receptor positive breast cancer [68]. Besides, aromatase inhibitors do not have

estrogenic effects on the liver and uterus, unlike tamoxifen [42]. Compared to tamoxifen in advanced disease, third generation aromatase inhibitors also show superiority in terms of clinical outcome, with different toxicities [48]. The Arimidex, Tamoxifen, Alone or in Combination (ATAC) clinical trial concluded that, after five years of treatment there is a significant improvement in disease free survival in patients treated with anastrozole (Arimidex®) alone and better outcome in ER-positive and PgR-negative patients. Another clinical trial, Breast International Group 1 – 98 (BIG1 – 98), which aimed to compare the efficacy of tamoxifen with letrozole (Femara®), concluded that, with two years of follow-up, there was a significant improvement in disease free survival with letrozole, comparing with tamoxifen. Additionally, a reduction of about 50% in the risk of contralesional breast cancer with letrozole, compared with tamoxifen, was also observed, as well as similar benefit in patients with ER-positive and PgR-negative breast cancer [35]. With five years of follow up, letrozole remains superior to tamoxifen, in terms of disease free survival [72]. Another clinical trial, Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability (TARGET), which aimed to compare efficacy and tolerability of anastrozole and tamoxifen in patients with advanced breast cancer, reported a clear superiority of anastrozole over tamoxifen, in terms of time for disease progression (Figure 7), clinical benefit and disease free survival, especially in hormone receptor positive cases [68].

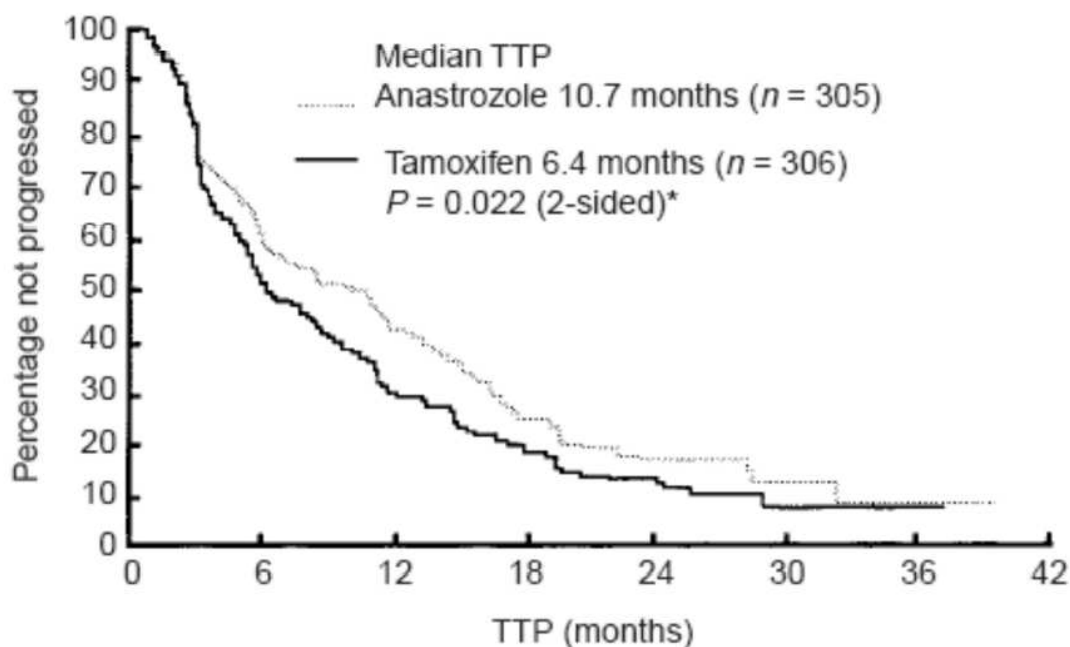


Figure 7: comparison of anastrozole and tamoxifen in terms of time for disease progression. Anastrozole is significantly superior to tamoxifen [70].

Concerning exemestane (*Aromasin*®), a clinical trial from the European Organization for Research and Treatment of Cancer (EORTC), which compared the efficacy of exemestane with tamoxifen, showed that the response rate is significantly higher with exemestane than with tamoxifen [69].

These drugs are administered orally on a daily basis [71] as hormonal therapy for ER-positive breast cancer in postmenopausal patients, since estrogens that are produced in the tumor and surrounding cells are important tumor growth stimulators in these cases. For this reason, today third generation aromatase inhibitors are used in the adjuvant setting for early breast cancer, to prevent recurrence, and in the neoadjuvant setting to reduce the size of the tumor [38]. Nowadays, these drugs are the most accepted treatment as an alternative to tamoxifen in the first line treatment for postmenopausal patients with advanced ER-positive breast cancer, due to their efficacy [72], as well as for metastatic breast cancer positive for hormone receptors in postmenopausal patients before and after surgery. In the neoadjuvant setting, aromatase inhibitors are used before surgery for patients with lymph node involvement and big tumors, since in these conditions, the possibility of presence of metastatic disease is considerable and this kind of treatment allows eliminating this possibility [35]. Nonetheless, all third generation aromatase inhibitors are indicated for second line treatment of postmenopausal breast cancer, in which progression after first line treatment with tamoxifen has occurred [68].

Today three types of selective aromatase inhibitors, which offer significant safety advantages over their non-selective ancestors, are available [69]. These aromatase inhibitors can be divided into two groups: steroidal and non-steroidal. Steroidal aromatase inhibitors bind irreversibly to aromatase and include exemestane [35]. These drugs are also known as inactivators, since they are reactive compounds that bind covalently to the enzyme's active site and irreversibly destroy its enzymatic activity [42]. Additionally, these drugs are androstenedione analogs that compete with endogenous androstenedione and testosterone for access to aromatase [72]. This activity requires that aromatase itself converts the inhibitor into a chemically reactive inhibitor which binds covalently and irreversibly to the protein structure of the enzyme-substrate binding site. Therefore, the enzymatic molecule is irreversibly inactivated and is no longer available to interact with other molecules. Steroid inhibitors have the potencial of having a great selectivity for the enzymatic target and long term efficacy, because the enzymatic activity restoration depends on the resynthesis of the enzyme itself and on the pharmacokinetics of the drug [39].

On the other hand, non-steroidal aromatase inhibitors bind reversibly to aromatase, and include anastrozole and letrozole [35]. These drugs are also called competitive inhibitors because they bind to the enzyme active site and block reversibly estradiol synthesis [42], being based on imidazole [72], namely the triazol functional group, which interacts with the aromatase heme prosthetic group [75]. All non-steroidal aromatase inhibitors have a nitrogen atom that allows them to interact with the iron atom on the heme group of aromatase. Their specificity for aromatase inhibition depends on structural aspects of the drug and they can reach a perfect fit into the substrate binding site. The better the fitting, the higher the affinity for the enzyme will be and also higher the limitation of the substrate binding to the enzyme [39]. On figure 8 the functional and structural differences between the three types of aromatase inhibitors are depicted.

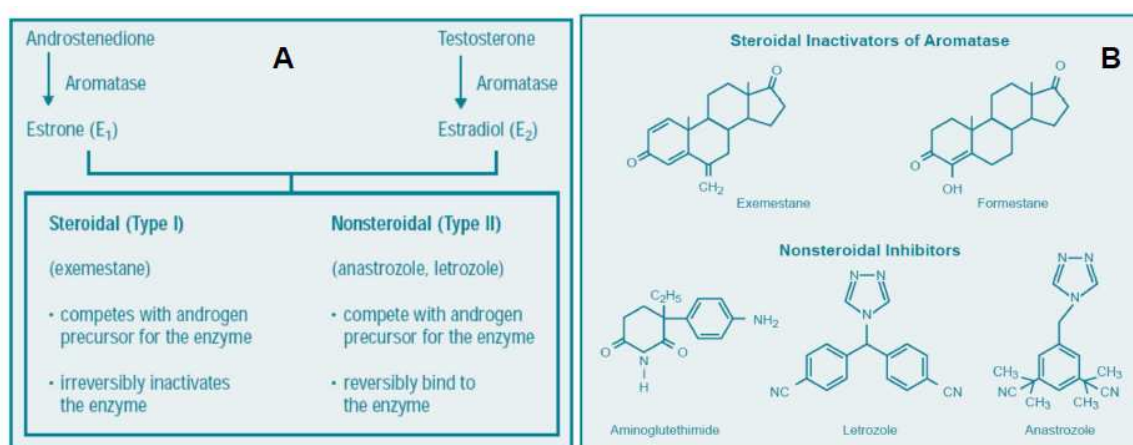


Figure 8: (A) differences between steroid and non-steroid aromatase inhibitors mechanisms of action. (B) Structural differences between all aromatase inhibitors described [71].

The molecular differences between the three third generation aromatase inhibitors used nowadays affect their selectivity for aromatase and therefore their ability to inhibit total body aromatization and endogenous estrogen suppression [72], and these pharmacokinetic and pharmacodynamic differences influence their choice, not only for advanced disease treatment, but also their potential use in the adjuvant setting. Additionally, there are also differences in their clinical pharmacology, as depicted in table 2.

Table 2: Differences between anastrozole, letrozole and exemestane in terms of daily dose, time to steady plasma levels, half-life and time to maximal estradiol suppression (Adapted from [70])

Feature	Aromatase Inhibitor		
	Anastrozole	Letrozole	Exemestane
Daily clinical dose (mg/day)	2	2.5	25
Time to steady-state plasma levels (days)	7	14-42	4
Half-life (hours)	40-50	48-96	24
Time to maximal estradiol suppression (days)	3-4	2-3	7

According to the data displayed on table 1 and 2 anastrozole is the most selective of the three types of third generation aromatase inhibitors. On the other hand, exesmetane is a steroid agent and therefore has an intrinsic antiaromatase activity, acting similarly to a weak androgen [68]. This drug is only used as second line treatment, after failure of other types of hormone therapy, while letrozole is administrated in both first and second line treatment for metastatic or locally advanced breast cancer, and anastrozole is more indicated for adjuvant treatment of early breast cancer, as well as first and second line treatment for metastatic ou locally advanced breast cancer [73].

Aromatase inhibitors are only used in postmenopausal women with ER- and/or PgR-positive breast cancer rather than in premenopausal women. The breast genetic promoter of aromatase is less sensitive to LH fluctuations than the one of the ovary, but is much more sensitive to increases in inflammatory cytokines, which are augmented in association with age and presence of proliferative breast disease. Aromatase inhibitors block estradiol production by the ovaries, breaking the negative feedback of LH and FSH inhibition. Besides, they can also obstruct cerebral aromatase as an alternative mechanism for the increase in FSH. This blocking can be counteracted in premenopausal women through the dramatic increase in androstenedione as a substrate, which is induced by the increase in LH. Additionally, the increase in FSH stimulates the production of aromatase itself. Thus, the ovary capacity to neutralize the increase in LH and FSH to overcome the blocking of estrogen production in premenopausal women renders aromatase inhibitors ineffective, when used by themselves. Therefore in premenopausal women, aromatase inhibitors are not used unless in special circumstances such as previous tamoxifen failure or medical contraindication for tamoxifen. In addition, in the case of being used in these women, aromatase inhibitors have to be combined with medical or surgical ovarian ablation [35].

3.2. Adverse Effects

In comparison with tamoxifen, aromatase inhibitors do not increase the risk of uterine cancers or thromboembolic events [35]. However, these drugs can have gynecological, musculoskeletal and cardiovascular adverse effects. This arises mainly due to their mechanism of action, which targets aromatase in a global way, resulting in a general estrogen deficient state [76], which can have an impact in sites where estrogen is necessary for normal function [37]. While tamoxifen is more associated to deep venous thrombosis and pulmonary embolisms than aromatase inhibitors, the frequency of nausea, hot flashes and gastrointestinal problems is comparable. Additionally, the use of aromatase inhibitors is also associated with an increase in osteopenia, osteoporosis, arthralgia and myalgia incidence [48]. Nonetheless, in a study comparing aromatase inhibitors with placebo, most of these problems were observed in both groups [77].

Gynecological effects of aromatase inhibitors can include vaginal dryness, loss of libido, painful intercourse and hot flashes [35]. In the ATAC trial, a slighter incidence of endometrial cancer was observed with anastrozole, compared to tamoxifen, as well as concerning hot flashes and weight gain [69].

Concerning musculoskeletal effects, these drugs can reduce bone density, increasing the risk of bone fractures [35]. This happens because estrogen is crucial in maintaining normal bone renewal and mass, and long term estrogen deprivation can be associated to osteoporosis and increase in bone fracture susceptibility. Low serum estradiol levels are associated with higher bone loss rate and increased risk of bone fractures in postmenopausal women. Besides, complete estrogen deprivation can result in an increase in the production of the cytokine receptor activator of nuclear factor kappa-B ligand (RANKL) by stromal cells and increased receptor activator of nuclear factor kappa-B (RANK) activity, resulting in an increase in the number of osteoclastic precursors and osteoclastogenesis. This, together with the reduction in osteoprotegerin circulating levels, results in an increase in the number of mature osteoclasts and, consequently, bone degradation [72]. Bisphosphonates can be used to prevent mineral loss caused by aromatase inhibitors [35].

Additionally, arthralgias and myalgias are also frequent during treatment with these drugs [35]. By definition, arthralgias consist in joint pain and stiffness, which, in patients treated with aromatase inhibitors are not associated with arthritis or inflammatory processes. The natural hypoestrogenemia associated with menopause is also associated with arthralgia, which can be treated with hormone replacement treatment or exacerbated with the use of aromatase inhibitors. Besides, it is also known

that estrogens regulate inflammatory cytokines and increase nociception (pain receptors) in the central nervous system [72]. So patients can experience pain in the hands, knees, hips, back, shoulders and/or feet, morning stiffness and difficulty falling asleep. The menopausal status seems to be an important risk factor for musculoskeletal effects in women, and the proportion of individuals with musculoskeletal pain increases with the increase in body mass index and decreases with physical activity. Other significant risk factors for arthralgia symptoms include the previous use of hormone replacement therapy, chemotherapy, geographic region, being more incident in the United States of America and less incident in the United Kingdom, body mass index higher than 30 and hormone receptor positivity. All of these factors are potentially related to a higher reduction in estrogen levels derived from hormone therapy [78].

The ATAC trial demonstrated a higher incidence in musculoskeletal problems and bone fractures with anastrozole compared with tamoxifen [69]. However, in the TARGET trial, no significant differences were observed in the frequency of bone fractures in women treated with anastrozole, compared with tamoxifen [68].

Despite aromatase inhibitors do not increase the risk of thromboembolic events, such as deep venous thrombosis, they result in an increase in ischemic cardiovascular events, due probably to estrogen depletion in the coronary arteries, resulting in a decrease in the vasodilatory response of estrogens to stress [35]. In fact, in the ATAC trial a smaller incidence in ischemic thrombotic cerebral events with anastrozole compared with tamoxifen was observed [69]. These data are supported by the TARGET trial, in which significantly less thromboembolic events were observed in women treated with anastrozole, compared with tamoxifen [68].

Additionally, high serum total cholesterol and low density lipoprotein-cholesterol (LDL-cholesterol) levels are important factors for the development of cardiovascular diseases, and low levels of high density lipoprotein-cholesterol (HDL-cholesterol) and hypertriglycemia are associated with coronary cardiac diseases and increase in morbidity and mortality. Estrogens have a protective effect in the human lipid profile and increased HDL levels and low LDL levels are associated with increased levels of estrogen. Therefore, estrogen level reduction, which occurs during aromatase inhibitor treatment can result in a more atherogenic lipid profile and increased risk of coronary disease [72].

3.3. Aromatase inhibitor resistance

It is known, since the establishment of hormone therapy use that some patients with hormone receptor positivity do not benefit from this kind of treatment. Additionally, even some patients who initially respond will eventually experience disease progression [71]. Even with initial benefit, women with ER positive metastatic breast cancer develop resistance and tumor recurrence. In most of these cases, the tumor remains ER positive [35].

Additionally, it is recognized that only the cases of ER or aromatase positive breast cancer respond to aromatase inhibitor treatment. Besides, an abnormally higher aromatase expression has already been demonstrated in the breast cancer cells and surrounding adipose stromal cells, when compared to the normal tissue. It is also believed that the *in situ* estrogen biosynthesis has a significant influence in tumor growth [75].

There are two types of endocrine resistance: *de novo* or intrinsic resistance, which consists in the absence of response upon initial exposure to the drug, and acquired resistance, which develops during the treatment, in patients who have initial response. Several aromatase inhibitor resistance mechanisms have already been suggested [75].

One possible mechanism consists in the upregulation of growth factors and/or associated signaling pathways, such as HER2, the epidermal growth factor receptor (EGFR) and the insulin-like growth factor receptor (IGFR) [35] and the crosstalk between all of these pathways [71]. In breast cancer patients, the overexpression or aberrant activation of *HER2* has already been demonstrated in breast tumors and associated to a worse prognosis with tamoxifen and endocrine resistance [79]. On the other hand, patients with *EGFR* and/or *HER2* overexpression and ER positive tumors respond significantly better to letrozole than to tamoxifen [80], since it has already been suggested that tamoxifen can act as a partial agonist in ER phosphorylation by EGFR or HER2 [81].

Conversely, PgR expression has already been demonstrated as being associated with greater benefit from tamoxifen [82]. The gene encoding PgR is an estrogen regulated gene and therefore is logical to think that PgR and ER positive tumors are truly dependent on estrogen. Thus, this kind of tumors has good response to treatment with tamoxifen, as well as with aromatase inhibitors. In the case of ER positive, PgR negative tumors, these can be less sensitive to estrogen or its nongenomic action [75]. Besides, it has been suggested that the activation of the phosphoinositide 3-kinase

/ protein kinase B (PI3K/AKT) pathway downregulates the transcription of the gene encoding PgR [83].

Concerning acquired resistance, the most obvious mechanism involves a selection process, since all tumors are heterogeneous [75] and tumor cells are highly plastic and capable of adapting to escape the normal mechanisms of cell proliferation control [48]. Each breast tumor contains ER positive and negative cells and during hormone treatment the population of ER positive cells decrease, and with time the ER-negative cells become the dominant cell type, and cells become resistant to hormonal therapy in this way [75].

However, clinically, many resistant tumors remain ER-positive. The resistance mechanisms in these cases are probably similar to those acknowledged for intrinsic resistance, except that resistance in these cases appears after treatment initiation. It is very unlikely that acquired resistance results from mutations in the aromatase or *ER* genes, originated during the treatment, but rather resistance results from a crosstalk between ER and growth factor pathways and other unidentified pathways. Due to the ability to adapt of breast cancer cells, hormonal therapies that involve the blocking of hormone dependent signaling cascades, necessary to their proliferation can originate new signaling mechanisms that overcome the effects of this kind of treatment. An interesting hypothesis is that resistance results from estrogen hypersensitivity or activation of its receptor independently of its presence. The increased expression of coactivators and decreased expression of corepressors associated to nuclear receptors has been associated to hypersensitive response of ER to estrogens [75]. In this case, studies in animal models have demonstrated that fulvestrant associated with letrozole is effective in the reduction of ER hypersensitivity, which causes its downregulation [84]. Another possible adaptation from breast cancer cells to low estrogen levels is that they can become insensitive to its antiapoptotic effects and, concomitantly respond in a paradoxal way to estradiol through mitochondrial and death receptor mediated apoptosis [75].

Currently several clinical trial are underway to overcome the problem of aromatase inhibitor resistance in breast cancer patients [48]. Based on the upregulation of growth factor pathways during acquired resistance, preclinical studies have demonstrated that some agents delay this resistance [85]. Therefore, several clinical trials have already been initiated, combining aromatase inhibitors with tyrosine kinase inhibitors or growth factor antagonists.

4. Pharmacogenomics

Pharmacogenomics is a new domain in the field of Oncology, which appeared in the turn of the century and allowed the discovery of numerous potential therapeutic targets and consequently new anticancer drugs. This area is based on the study of the function of the genome, especially from the tumor, drug activity, allowing the development of new tumor selective drugs and the adaptation of the treatment to the genetic features of each tumor [44].

Different patients respond in different ways to the same treatment, and these differences are frequently higher in members of one population, than in the same individual in different time points. This is consistent with the heritage of drug response determinants, and it is estimated that genetics contributes to 20 to 95% of the variability in drug disposition and effects. However, several non-genetic factors contribute also to the different effects of drugs, such as, age, organ function, concomitant therapies, interactions with other drugs and the nature of the disease. Nonetheless, currently there are several examples in which drug response variability is derived from variants in the sequence of genes encoding metabolizing, transporter and drug target enzymes. Pharmacogenomics applies strategies in a genomic level to evaluate the hereditary basis of these differences in drug response, with the advantage that these hereditary determinants remain stable throughout life [86].

4.1. Polymorphisms

Polymorphisms consist of alterations in the DNA sequences which occur in more than 1% of the population, resulting in different allelic forms of one gene, and are hereditary and their configuration is unique of each individual [38].

Genetic polymorphisms have been largely studied in Oncology, and the risk of cancer appears to be strongly dependent on the genetic profile of the individuals exposed to the same carcinogenic environment. This kind of studies is defined as Molecular Epidemiology, a field that was originated in the 1990s and is rapidly growing [44]. More than 1,4 million polymorphisms involving a single nucleotide – Single Nucleotide Polymorphisms (SNPs) – have already been identified in the human genome, and 60 thousand of these are located in coding regions. These polymorphisms can affect drug metabolizing, transporter and target enzymes. There are over 30 families of drug metabolizing enzymes in humans, and almost all of them have genetic variants, which can result in proteins with functional alterations. On the other hand, drug transporter

enzymes have an important role in the regulation of absorption, distribution and excretion of several drugs, and it is logical to assume that variants in these proteins can have important consequences in drug response. Finally, genetic variants in the drug targets can also have important effects in their efficacy [86].

Concerning the study of polymorphisms in genes encoding these kinds of enzymes, the haplotype study, that is, the combined effect of several polymorphisms in the same individual is a better response predictor than individual polymorphisms [86]. Besides, the importance of linkage disequilibrium (non-random association of alleles at two or more loci, not necessarily on the same chromosome) and the data from haplotype studies for application in association studies is increasingly acknowledged [87]. In fact, due to the weak association already observed in SNP based studies, the haplotype based studies have been proposed as more powerful for the identification of genetic variants causative of complex genetic diseases such as cancer [32].

Studies have demonstrated that the human genome is composed of genomic segments (blocks) which can have little evidence of historic recombination and low haplotype diversity [88]. A study design frequently employed to analyze this type of genes is the association study, in which the frequency of polymorphic genotypes are compared between groups with different phenotypes, with the goal of analyzing polymorphisms putatively related with disease risk or in linkage disequilibrium with causative variants. More simple polymorphisms are the biallelic, which usually arise from one SNP and originate three different genotypes: homozygote for the most common or major allele, homozygote for the rarest or minor allele and heterozygote. Multiallelic repetition or microsatellite polymorphisms can also be studied, but originate more different genotypes. Genetic association studies based on populations are more powerful than linkage or pedigree studies for low penetrance allele identification, which by definition may not be expressed in different members of the same family [89].

Therefore, the potential of Pharmacogenomics is huge, giving rise to an array of molecular diagnosis methods, which can become routine tools that allow selecting the most adequate medication and dose for each patient individually. Besides, the genotype of a patient only has to be determined once for each gene for it remains unchanged throughout life, unless a rare somatic mutation occurs [86].

4.2. CYP19A1 polymorphisms

Considering that hormones play a crucial role in the development and progression of breast cancer, polymorphisms in genes that regulate endogenous

estrogen levels are attractive candidates for risk factors for breast cancer [90]. Therefore, it is biologically plausible that polymorphisms in the *CYP19A1* gene, which encodes aromatase, which in turn synthesizes estrogen, may be related to breast cancer risk, survival and response to treatment [91]. Additionally, it is known that mutations in key breast cancer susceptibility genes, such as *BRCA1* and *BRCA2*, which are highly penetrant, only explain a small proportion of breast cancer cases in the general population (less than 5%), but the genetic risk factors that contribute to sporadic breast cancer are still largely unknown [18], and *CYP19A1* polymorphisms may, in part have a role in breast cancer risk. In addition, it has already been suggested that *CYP19A1* polymorphisms may affect the circulating estrogen levels in postmenopausal women and, in this way, affect breast cancer risk [43, 45, 92, 93]. Although there are many epidemiological studies of common *CYP19A1* polymorphisms, the results concerning breast cancer risk are inconsistent [9].

Additionally, the response to the treatment with aromatase inhibitors, directed specifically to the aromatase enzyme, can vary dramatically, ranging from lack of efficacy to severe adverse effects. A possible explanation for this variability may be variation in the gene that encodes the drug target, *CYP19A1*, or even drug metabolizing enzymes [94], hence the importance of the study of polymorphisms in the *CYP19A1* gene, not only for breast cancer risk assessment, but also for the evaluation of drug response. On the other hand, many somatic genetic alterations correlate to breast cancer prognosis and survival, but little is known about the influence of common germ line variations in the clinical outcome [95].

More than 80 *CYP19A1* gene polymorphisms have already been identified, being the majority SNPs present in non-coding regions or synonymous in coding regions, that is, in which the polymorphism does not affect the protein structure [38].

Some variations in the *CYP19A1* gene are associated with altered sex hormone levels in the serum or urine, providing an explanation for the increased breast cancer risk in relation to estrogen exposure [29]. However, previous population studies of common *CYP19A1* polymorphisms have originated inconsistent results, concerning its potential association with sex hormone levels or risk of estrogen dependent conditions [87]. Nonetheless, *CYP19A1* polymorphisms have already been demonstrated as potentially associated with differences in the gene expression and risk of breast and prostate cancer. However, its role in aromatase inhibitor response has not yet been fully explored [44], but the studies already published suggest that alterations in the gene may result in a phenotype functionally less sensitive to this therapy [96].

Besides, polymorphisms in the *CYP19A1* gene may be involved in the regulation of the enzymatic activity through mRNA stabilization mechanisms, increase in transcription or post-translational regulation of expression [2]. Additionally, it has already been demonstrated that *CYP19A1* polymorphisms alter aromatase activity in postmenopausal women [45]. Therefore, although many *CYP19A1* polymorphisms have already been identified their possible functional importance remains poorly understood [87].

Being the *CYP19A1* gene so complex, its polymorphisms may have various locations, as evident in figure 9.

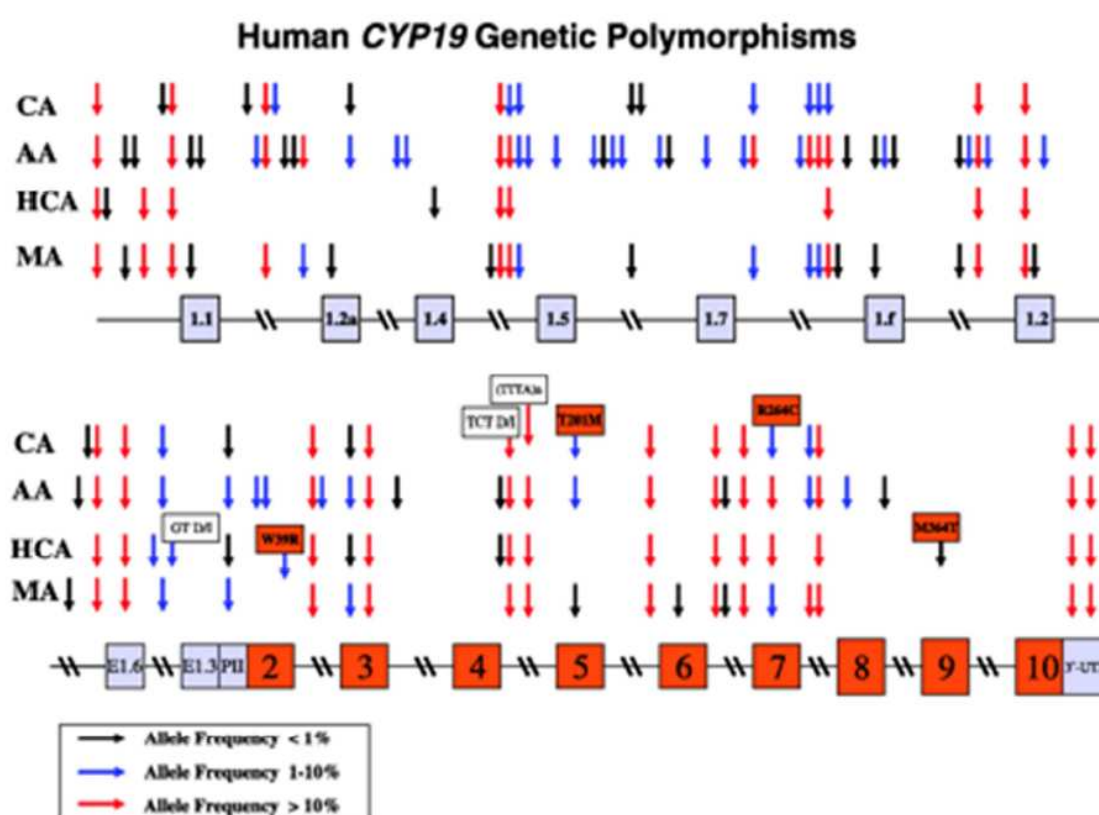
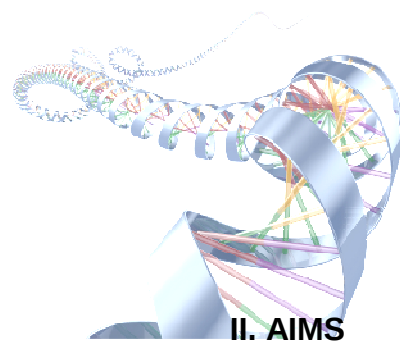


Figure 9: Localization and allelic frequency of the already identified *CYP19A1* polymorphisms [87].

As observed in the figure, most of the already identified polymorphisms have an intronic localization. In fact, there are only four non-synonymous polymorphisms identified in coding regions: Trp³⁹Arg and Thr²⁰Met, which do not affect aromatase activity [87], and Arg²⁶⁴Cys and Met³⁶⁴Thr, which reduce enzymatic activity [38].

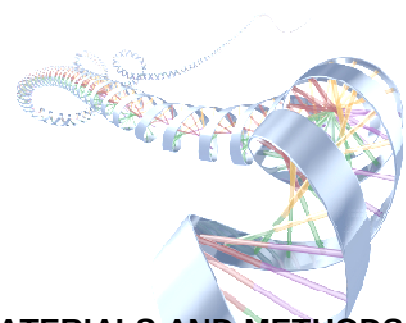
Although the results of different studies based on *CYP19A1* polymorphisms and its effects are very controversial, all of them suggest an important role for genetic variation in the aromatase function and, as a result, estrogen production [94].



The general aim of this study was to study the influence of genetic polymorphisms in the aromatase expressing gene in the response of breast cancer patients treated with aromatase inhibitors.

Specific aims were the following:

- To review the literature for candidate polymorphisms that could have an effect on the response to aromatase inhibitors;
- To characterize the sample population and seek for an overrepresentation of any of the studied alleles in any subset;
- To perform a survival analysis to assess whether any of the alleles has an influence; and
- To perform linkage disequilibrium analysis to evaluate if there is a linkage relationship between the studied alleles.



II. MATERIALS AND METHODS

1. Population

In the present study, biological samples were obtained from breast cancer patients treated with the aromatase inhibitors letrozole and/or exemestane in the Portuguese Institute of Oncology Francisco Gentil of Porto (IPOP), between 1989 and 2007, together with an informed consent form of participation in the project.

For patient selection, inclusion criteria used were women with pathologically confirmed invasive breast cancer, subjected to aromatase inhibitor treatment as a second line treatment after tamoxifen or as palliative treatment after disease progression, without previous anti-neoplastic treatment or previous neoplasms, ER positivity, survival of more than one year, followed in the mentioned institution and with a clinical file and informed consent form of participation in the study. Exclusion criteria were considered as presence of other neoplasms, absence of pathological diagnosis, non-invasive breast cancer, absence of ER status determination, absence of clinical file and indication that the patients did not have capacity to understand the study – absence of consent form.

After this selection, we ended up with 109 patients, from which we had access to the clinical files in order to collect the necessary data for this study.

2. Sample Processing

For this study, 8ml of peripheral blood were collected from participant patients, through a standardized method of intravenous collection to sterile tubes containing a solution of ethylenediaminetetraacetic acid (EDTA). Two complete blood count (CBC) tubes were therefore obtained.

All the tubes were centrifuged for approximately 10 minutes at 2500 rotations per minute (rpm). In this way, in the CBC tubes, the blood components, such as erythrocytes, leucocytes and platelets, were separated from the plasma.

Subsequently, 1ml of plasma, 1ml and 200µl of total blood from CBC tubes were obtained and transferred to previously identified eppendorf tubes. The remaining blood from the CBC tubes was transferred into a 50ml Falcon tube, for further processing.

To this Falcon tube a hypotonic solution (AKE 1x) was added making up 50ml, to lysate the erythrocytes, and incubated at 4°C for 30 minutes. After this, the tubes were homogenized and centrifuged for 10 minutes at 2500 rpm. The

supernatant was discarded and the pellet resuspended with AKE 1x, with another centrifugation for 10 minutes at 2500 rpm. The supernatant was again discarded and the pellet resuspended, this time in 30ml of phosphate buffered saline (PBS) 1x, to wash the cells. The solution was again centrifuged for 10 minutes at 2500 rpm, and the supernatant rejected and pellet resuspended in the remaining PBS.

Finally, aliquots were made from this pellet to identified eppendorfs and saved at -80°C, or homogenized in 2ml of Tripure® isolation reagent (Roche™), with a syringe and needle, for the cells to be well separated, and transferred to two previously identified eppendorfs and saved at -80°C.

2.1. Genomic DNA isolation

Genomic DNA isolation was performed using the QIAamp® DNA Mini Kit (Qiagen™), from peripheral blood nucleated cells, or cell pellet, according to the manufacturer protocol, obtaining 200µl of DNA, which was diluted in a 5:50 proportion and stored at 4°C for use, with the undiluted DNA being stored at -40°C.

3. Genotyping

3.1. Polymorphism Selection

The analyzed polymorphisms were selected based on a review from the literature through the National Center for Biotechnology Information (NCBI) website (www.pubmed.com). We have selected five polymorphisms, from which four are SNPs and one is a variable number tandem repeat (VNTR) polymorphism. Two of the SNPs – rs4646 (G/A substitution) and rs10046 (A/G substitution) – are located in the 3'UTR (untranslated region) of the *CYP19A1* gene, one – rs700518 (C/T substitution) – is located in exon 3, being a synonymous polymorphism, and one – rs6493497 (G/A substitution) – is located in the 5'-flanking region (5'FR) of exon 1.1. The VNTR polymorphism is a tetranucleotide (TTTA)_n repetition polymorphism and is located in intron 4 of the *CYP19A1* gene. These polymorphisms were selected on the basis of their gene location and minor allelic frequency in Caucasians, being more than 10%.

3.2. Genotyping

The genotypic analysis of the rs10046, rs4646, rs700518 and rs6493497 SNPs was performed by Real-Time Polymerase Chain Reaction (PCR), using the

7300 Real Time PCR System™ from Applied Biosystems®. For the reaction mix, Taqman™ Genotyping Master Mix (Applied Biosystems®) was added to the Taqman™ Drug Metabolism Genotyping assays (Applied Biosystems®, table 4) and to double distilled water, in the following quantities, according to the concentration in which the assays were purchased (20X or 40X):

Table 3: Mix conditions for the two possible assay dilutions (20x and 40x)

Assay Dilution	20X	40X
Double distilled water	2,25	2,375
Master Mix	2,5	2,5
Assay	0,25	0,125

For the rs10046, rs4646, rs700518 and rs6493497 SNPs, the assays had the following references and concentrations:

Table 4: Assay references and dilutions for the four SNPs studied

SNP	Reference	Concentration
Rs10046	C__8234731_30	20X
Rs4646	C__8234730_1_	20X
Rs700518	C__8794675_30	20X
Rs6493497	C__29374681_10	40X

The allelic discrimination was analyzed using the 7300 System SDS™ Software v1.2.3 from Applied Biosystems®.

3.3. DNA Sequencing

To analyze the tetranucleotide repetition polymorphism (TTTA)_n, the samples were sequenced in the region where it is located, in intron 4.

3.3.1. DNA Amplification

To prepare the samples for the sequencing, they were amplified by PCR to obtain a 312bp fragment. The PCR reaction was performed using two different protocols, based on the enzyme *Taq* DNA polymerase used.

On the first protocol, 0,2µg of DNA were amplified in a PCR mix with a total volume of 50µl. To prepare the PCR mix, 1 unit of *Taq* DNA Polymerase, 1x reaction buffer, and 2,5 mM of MgCl₂ (all from Amersham – GE LifeSciences®) were used, as well as 0,30mM of deoxiribonucleotides (dNTPs), 15 µM of each primer (Forward: 5' –

GAT AGA GTC AGA GCC TGT CTC – 3' and reverse: 5' – CAC CAA AGT TTC CTG TCT TTT CAG – 3') and 34,8 µl of double distilled water.

On the other hand, on the second protocol, the PCR mix was also 50 µl, using 0,2 µg of DNA. In this protocol, however, a different polymerase was used: *DreamTaq™* Green DNA Polymerase (Fermentas®), which is accompanied by a PCR reaction buffer (*DreamTaq™* Green Buffer) which already contains MgCl₂. So, to maintain the already optimized PCR conditions it was necessary to perform a compensation in the amount of buffer used so that a sufficient volume is added to contain the required amount of magnesium. Therefore to the DNA, one unit of *DreamTaq™* Green DNA polymerase, 6,25 µl of *DreamTaq™* Green Buffer, 0,3 mM of dNTPs, 0,15mM of each primer and 38,55 µl of double distilled water were added.

In both protocols, the PCR reaction conditions were the same, being performed in the *MyCycler™* Thermal Cycler (Bio-Rad®). The reaction included an initial denaturation step at 95°C for five minutes, followed by 35 cycles of denaturation at 94°C for 1 minute, annealing at 55°C for 1 minute and extension at 72°C for 1 minute, ending with a final extension at 72°C for 4 minutes.

3.3.2. DNA Amplification Validation

Next, the PCR products obtained were subjected to agarose gel electrophoresis to confirm their proper amplification. The electrophoresis was performed in 1,5% agarose gels, using GRS Agarose LE™ (Grisp®) in TBE (Tris/Borate/EDTA) buffer, and stained with 10µg/ml of ethidium bromide. The gel was loaded with 13µl of PCR product, after adding 6x Loading Dye™ from Fermentas®. The O'GeneRuler™ 100bp DNA ladder (Fermentas®) was also loaded to assess the correct PCR product size. The electrophoresis was performed on a Power Pac Basic™ electrophoretic system (Bio-Rad®), at 130 volts and 300milliAmperes for 20 to 30 minutes. Finally, the gels were visualized in a Bio-Rad® transilluminator, and analyzed using the Quantity One™ Gel Doc XR v4.65 software from Bio-Rad®.

3.3.3. Sequencing

For the determination of the sequence of the amplified region, the PCR products were sent to the Institute of Molecular Pathology and Immunology of the University of Porto, who kindly offered their collaboration in this study.

The samples were purified by adding 2 µl of the Exo-SAP-IT™ (Affimetrix®) purification reagent to 5µl of PCR product and incubating at 37°C for 15 minutes to

degrade the residual primers and nucleotides. Subsequently, the samples were incubated at 80°C for another 15 minutes to inactivate the purification reagent.

Next, the sequencing reaction was performed using 1,5 µl of purified PCR product (2,5 µl were used in cases where the amount of product was lower), 0,4 µl of BigDye™ Terminator v3.1 Cycle (Applied Biosystems®) sequencing kit, 0,5 µl of BigDye™ Terminator v1.1, v3.1 5x (Applied Biosystems®), 0,4 µl of the reverse primer used in the DNA amplification (100mM) and 2,2 µl of double distilled water.

The samples were sequenced under the following conditions: 2 minutes at 96°C, followed by 30 cycles of 10 seconds at 95°C, 10 seconds at 55°C and 140 seconds at 70°C, and finished with 10 minutes at 60°C.

After that, the products were purified again, using the Illustra Sephadex™ G-50 Fine DNA Grade (Sigma-Aldrich®) and sequenced using the 3130xl Genetic Analyzer™ (Applied Biosystems®) instrument. And finally, the samples were analyzed using the Sequencing Analysis™ v5.2 (Applied Biosystems®) software.

Finally, for the analysis and comparison of the sequences and localization of the (TTTA)_n polymorphism, as well as the TCT insertion/deletion 50bp upstream of the first repetition, the Sequence Scanner™ v1.0 (Applied Biosystems®) software was used.

4. Statistical analysis of the results

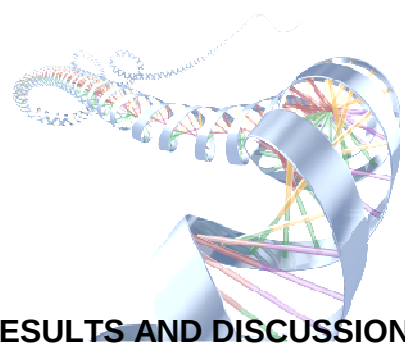
Comparison of the allelic frequencies obtained in our study with the ones reported in the literature was performed using the Chi-Square Test, with a 2x2 contingency table, and a 5% significance level in which if the *p*-value was higher than 0.050, the results are in accordance with each other. These results were obtained using the *Epi Info*® Software (version 3.3, 2004).

Characterization of the allelic distributions stratified for different patient, clinical and tumor parameters was assessed also using a Chi-Square Test with a 5% significance level, because all the variables were categorical, in which if the *p*-value obtained was lower than 0.050, parameters were not independent and related with each other. This analysis was performed using the SPSS® Statistics Software (version 19, 2010).

Survival analysis was obtained using the Kaplan-Meier Method, with significance calculated through the log-rank (Mantel-Cox) test, in which if the *p*-value was lower than 0.050, there are significant differences between parameters. The time variable calculated in these analyses was defined as time between diagnosis and

death, in the case of overall survival, or progression, in the case of progression free survival. This analysis was also performed using the SPSS® Statistics Software (version 19, 2010).

Finally, linkage disequilibrium relationships were assessed using the Chi-square Test for association, with 3x3 contingency tables and in which if the *p*-value was lower than 0.050, the alleles are associated significantly. This was performed using the *EH*® program for linkage disequilibrium estimation.



III. RESULTS AND DISCUSSION

Over the past decades Pharmacogenomics has become more and more important in Oncology, not only in determining new risk factors for cancer in Molecular Epidemiology, but also for assessment of predictors of response to treatment so it can be more individualized and one can reduce the risk of adverse effects or even absence of response to treatments.

Being estrogen such an important factor for breast cancer development and progression, it is plausible to assume that the mechanisms of its synthesis or degradation are very important in this field. And also, being aromatase one of the molecules of estrogen biosynthesis, the importance of its study is reinforced.

Although there is a large number of studies concerning the effects of the *CYP19A1* polymorphisms on breast cancer risk, studies concerning their effects on treatment response are very scarce. And this is an important issue, because aromatase is the target of a relatively new breast cancer treatment modality, which is aromatase inhibitors that have already been demonstrated to be very effective. However, as for almost all drugs not all the patients respond to aromatase inhibitors in the same way and some do not even respond or have serious side effects. So it would be logical that its target could have a causal effect in these differences and that is what we are aiming to assess in this study. The main objectives of this work are to analyze some polymorphisms in the *CYP19A1* gene and try to find any effect on treatment response. So this study is divided in essentially five parts: the first one consists of a review of the available literature for candidate polymorphism selection, next we analyzed the allelic distributions of the selected polymorphisms and compared them to the ones reported by other studies, after that the study population was characterized and we sought for associations between the population characteristics and the different polymorphisms. Afterwards, we analyzed the effects of the polymorphisms on aromatase inhibitor treatment response, in terms of overall and progression free survival. Finally, we assessed the relationship of the polymorphisms with each other in terms of linkage disequilibrium. All of these results will be discussed next.

1. Studied Polymorphisms

1.1. *Polymorphism Selection*

First of all, we performed a review of the literature to find which are the most interesting candidates in this context, namely the polymorphisms that would be more likely to have an effect on aromatase inhibitor treatment response.

Probably the most studied polymorphism in the context of breast cancer is the tetranucleotide repetition polymorphism (TTTA)_n, located in intron 4 and a three nucleotide (TCT) deletion, located 50 base pairs upstream of it. This polymorphism has already been extensively studied in relation to various factors associated to breast cancer, such as risk, response to treatment, survival, among others [28, 29, 31, 32, 38, 45, 91, 93, 95-100]. Other largely studied polymorphisms in relation to breast cancer, both located in the 3'UTR of exon 10 are rs10046 [29, 45, 91, 92, 96, 101-104] and rs4646 [29, 96, 105]. However the majority of these studies have conflicting results, and there is still no consensus as to the effect of these polymorphisms in breast risk, survival and response to treatment.

Therefore, those three were included in this study. Also we wanted to test other polymorphisms with other locations in the gene so our study could be more comprehensive. Another condition, and since our sample population was not very large (109 patients), was that the polymorphisms should have a reported allelic frequency relatively high. Our threshold was established for at least 10% in Caucasians, reported in the literature. So, we ended up with five polymorphisms: one VNTR, the tetranucleotide repetition polymorphism, and four SNPs. In terms of gene localization, the tetranucleotide repetition polymorphism is intronic (intron 4), both rs4646 and rs10046 are located in the 3'UTR of exon 10, rs700518 has an exonic localization, resulting in a synonymous amino acid change and finally, rs6493497 is located in the 5'Flanking Region of exon 1.1. Also all of these polymorphisms have a reported allelic frequency higher than 10%.

1.2. *Allelic and Genotypic Frequencies Observed*

After the selection of the polymorphisms that we were going to analyze, we determined the allelic frequencies of each one, and also compared them with the ones reported on the literature based on Caucasian populations.

The results for the SNPs are shown in table 5. For rs4646, the minor allele is present in 24.8% of the studied population, while rs10046 and rs700518 have a similar distribution with a 48.2% minor allele frequency, which is very frequent in our population. On the other hand, rs6493497 is the rarest SNP, with a minor allele

frequency of 15.6%. Additionally, our results are consistent with the ones reported in the literature ($p>0.050$), as shown in table 5

Table 5: Allelic frequencies obtained in our study and comparison to the ones reported in the literature

Population	Minor Allele n(%)	Major Allele n(%)	p^a	Reference
rs4646				
Portuguese	54 (24.7)	164 (75.2)	Ref.	Schmid, 2011
Spanish	43 (32.1)	91 (67.9)	0.136	[112]
American	139 (28.9)	341 (71.1)	0.252	[96]
Spanish	46 (24.2)	144 (75.8)	0.896	[105]
American	759 (28.0)	1951 (72.0)	0.305	[20]
American	435 (29.0)	1065 (71.0)	0.196	[11]
American	630 (25.0)	1892 (75.0)	0.945	[115]
German	674 (30.4)	1546 (69.6)	0.085	[29]
rs10046				
Portuguese	105 (48.2)	113 (51.8)	Ref.	Schmid, 2011
Spanish	74 (55.2)	60 (44.8)	0.198	[112]
American	211 (44.0)	269 (56.0)	0.301	[96]
Spanish	87 (45.8)	103 (54.2)	0.632	[105]
American	1192 (44.0)	1518 (56.0)	0.232	[20]
American	660 (44.0)	840 (56.0)	0.248	[11]
German	1330 (52.2)	1218 (47.8)	0.253	[29]
Norwegian	430 (44.7)	532 (55.3)	0.353	[92]
United Kingdom	733 (48.0)	794 (52.0)	0.964	[95]
United Kingdom	1833 (52.4)	1661 (47.5)	0.218	[45]
rs700518				
Portuguese	105 (48.2)	113 (51.8)	Ref.	Schmid, 2011
American	221 (46.0)	259 (54.0)	0.602	[96]
American	1247 (46.0)	1463 (54.0)	0.540	[20]
German	386 (49.2)	398 (50.8)	0.780	[113]
American	1048 (50.0)	1048 (50.0)	0.606	[114]
American	1211 (48.0)	1311 (52.0)	0.967	[115]
rs6493497				
Portuguese	28 (13.2)	184 (86.8)	Ref.	Schmid, 2011
American	77 (16.0)	403 (84.0)	0.338	[96]
American	240 (16.0)	1260 (84.0)	0.295	[11]

^a Chi-square test for comparison of our results to the ones reported in the literature, in which if $p>0.050$ the variables are not statistically different.

On the other hand, concerning the tetranucleotide repetition polymorphism, the results are summarized in table 6. We observed that the most frequent alleles in our population are the seven repetition with a TCT deletion (39.8%), followed by the eleven repetition allele (25.3%), the seven repetition allele (11.9%) and the eight repetition allele (11.8%). All of these alleles were for that reason included in the subsequent analyses performed in this work.

Our results for the allelic frequencies of the different alleles of the tetranucleotide repetition polymorphism are in general consistent with the ones reported in the literature. However, concerning the nine repetition allele, our results were inconsistent with all the papers reviewed, as well as other alleles with some studies. Concerning the nine repetition allele, we could not find any intrinsic factor in the populations included in the other studies that could explain these differences. We can assume that possibly this allele may be a marker of the Portuguese population, since in our population its frequency was consistently higher than the ones reported in the other studies. However, this issue needs more detailed analysis, to find out the true origin of these differences. Additionally, the differences may be due to the large inconsistency concerning this polymorphism. For example, many studies do not consider the TCT deletion associated with the seven repetition allele as a separate allele [28, 93], as we did. Besides, some studies report the presence of an allele containing six repetitions [28] and others the presence of a thirteen repetition allele [28, 43, 93, 97]. However, these alleles referred in those studies are all relatively rare, and that may be the reason why we did not find them in our population.

Table 6: Allelic frequencies obtained in our study for the tetranucleotide repetition polymorphism and comparison to the ones reported in the literature

Population	Allele Present n(%)	Allele Absent n(%)	p^a	Reference
(TTTA)₇				
Portuguese	22 (11.9)	164 (88.1)	Ref.	Schmid, 2011
United Kingdom	169 (14.1)	1029 (85.9)	0.402	[45]
United Kingdom	84 (12.8)	570 (87.2)	0.713	[3]
Russian	121 (16.5)	611 (83.5)	0.114	[4]
(TTTA)_{7(-3bp)}				
Portuguese	74 (39.8)	112 (60.2)	Ref.	Schmid, 2011
United Kingdom	405 (33.8)	793 (66.2)	0.111	[45]
United Kingdom	209 (32.0)	445 (68.0)	0.046	[3]
Russian	199 (27.1)	533 (72.8)	0.001	[4]
United Kingdom	639 (36.7)	1103 (63.3)	0.405	[47]

Table 6 (continued): Allelic frequencies obtained in our study for the tetranucleotide repetition polymorphism and comparison to the ones reported in the literature

Population	Allele Present n(%)	Allele Absent n(%)	p^a	Reference
(TTTA)₈				
Portuguese	22 (11.8)	112 (60.9)	Ref.	Schmid, 2011
Brazilian	45 (8.8)	511 (91.2)	0.003	[32]
United Kingdom	107 (8.9)	1091 (91.1)	0.005	[45]
United Kingdom	88 (13.5)	566 (86.5)	0.367	[3]
Russian	119 (16.3)	613 (83.7)	0.963	[4]
American	95 (10.3)	829 (89.7)	0.034	[102]
(TTTA)₉				
Portuguese	14 (7.5)	172 (92.5)	Ref.	Schmid, 2011
Brazilian	2 (0.4)	509 (99.6)	<0.001	[32]
United Kingdom	6 (0.5)	1192 (99.5)	<0.001	[45]
United Kingdom	7 (1.1)	647 (98.9)	<0.001	[3]
Russian	1 (0.1)	731 (99.9)	<0.001	[4]
American	2 (0.2)	922 (99.8)	<0.001	[102]
(TTTA)₁₀				
Portuguese	4 (2.2)	182 (97.8)	Ref.	Schmid, 2011
Brazilian	5 (1.0)	506 (98.9)	0.225	[32]
United Kingdom	15 (1.3)	1183 (98.7)	0.327	[45]
United Kingdom	10 (1.5)	644 (98.5)	0.559	[3]
Russian	15 (2.0)	715 (98.0)	0.935	[4]
American	21 (2.3)	903 (97.7)	0.918	[102]
(TTTA)₁₁				
Portuguese	47 (25.3)	139 (74.7)	Ref.	Schmid, 2011
Brazilian	160 (31.1)	351 (68.9)	0.123	[32]
United Kingdom	458 (38.2)	740 (61.8)	0.001	[45]
United Kingdom	237 (36.2)	417 (63.8)	0.005	[3]
Russian	243 (33.2)	489 (66.8)	0.038	[4]
American	312 (33.8)	612 (66.2)	0.024	[102]
(TTTA)₁₂				
Portuguese	3 (1.6)	183 (98.4)	Ref.	Schmid, 2011
Brazilian	16 (3.1)	495 (96.9)	0.276	[32]
United Kingdom	31 (2.6)	1167 (97.4)	0.424	[45]
United Kingdom	15 (2.3)	639 (97.7)	0.572	[3]
Russian	34 (4.6)	698 (95.4)	0.060	[4]
American	29 (3.1)	895 (96.9)	0.257	[102]

^a Chi-square test for comparison of our results to the ones reported in the literature, in which if $p > 0.050$ the variables are not statistically different.

2. Population Characterization

In this study a total of 109 breast cancer patients were who were treated with letrozole and/or exemestane either as a first line treatment or as palliative therapy after recurrence from tamoxifen. From these patients we collected data from their clinical files in order to characterize the study population and assess if there is any of the studied alleles is overrepresented in any of the subsets.

We collected patient characteristics and tumor features, whose distributions are shown in table 7. From the results we can observe that the median age at diagnosis is 55 years of age, with most of the patients (61.5%) being diagnosed after 50 years of age. Also, most of the patients were postmenopausal at diagnosis (64.9%). This is logical because hormonal therapies and especially aromatase inhibitors are only administered to postmenopausal women, because they are not effective on premenopausal patients [35]. Also, the patients who were premenopausal at diagnosis, became postmenopausal during the course of the treatment, or else they would not be administered aromatase inhibitors. Considering age at menarche, we divided this in early and late, assuming 13 years of age as the cutoff. We observed that the majority of the patients from which we obtained this information had menarche after 13 years of age (60.2% as opposed to 39.8% which had early menarche). And as to age at menopause, which was divided in early and late for before and after 50 years of age, respectively, most of the patients (60.4%) had an early menopause, compared to 39.6% of the patients which had late menopause.

Concerning family history of breast cancer, the majority of the patients included in our study (66.3%) had no previous history of breast cancer in their family. Finally, concerning body mass index, which we considered healthy if less than 30kg/m² and obese above that value, most of the patients had an healthy weight (69.4%).

So, in summary, our study population is composed mostly of women over the age of 50, postmenopausal, with late menarche and early menopause, no family history of breast cancer and not obese. Therefore, the majority of the patients had no increased estrogen exposure or increased risk of having breast cancer.

We also assessed the tumor characteristics in our clinical file review. We evaluated the presence of an *in situ* component associated with the primary tumor, in which only 32.1% of the patients had this feature. All of the tumors were invasive, with most of them being ductal carcinomas (91.0%), and the others not ductal were lobular, papillary or mucinous carcinomas. Also, most of the tumors were low to intermediate in histological grade (71.6%).

Concerning the hormonal status, all of the tumors were ER-positive and most of them were also PgR-positive (82.0%). We only obtained information about the HER2 status for a small subset of the patients (36 out of 109), because this information was only collected by the doctors in cases in which patients needed treatment with Trastuzumab (Herceptin®), or more advanced cases, in which there was failure of palliative hormone therapy. Nonetheless, from this subset of patients, 55.6% were HER2-positive.

Concerning the tumor staging, we did this analysis on the basis of the American Joint Committee on Cancer (AJCC) TNM (Tumor size, lymph Nodes and distant Metastases) system for pathological staging after surgery. Concerning tumor size, we considered tumors smaller than 20mm as small, and larger than 20mm as large. For the lymph node involvement we considered it as being high in the presence of more than 10 nodes metastasized (N_3) and low when less than 9 nodes were metastasized (N_0 , N_1 and N_2). Finally, for presence of distant metastases we divided the population in with or without distant metastases. Therefore, most of the patients had large tumors (57.4), low lymph node involvement (75.7%) and no distant metastases (93.5%). We also assessed these data altogether and divided the cases into early stage breast cancer (stage I and II) and advanced stage breast cancer (stages III and IV). Most of the patients had early stage breast cancer (68.9%).

In summary, the studied population is composed mostly of patients bearing ductal invasive carcinomas, with no *in situ* component associated, of low to intermediate histological grade, ER-, PgR- and HER2-positive. In addition, most of the tumors were large but with low lymph node metastatic involvement and no distant metastases, being early stage breast cancer cases.

Table 7: Overall distributions of clinicopathological and tumor features collected from the patients' medical files

Parameter	n	%
Age at diagnosis (y)	Median age: 55 (44.34 – 65.66)	
>50	71	61.5
<50	38	34.9
Total known	109	100.0
Menopausal status		
Postmenopausal	61	64.9
Premenopausal	33	35.1
Total known	94	100.0
Age at menarche (y)	Median age: 13 (11.21 – 14.79)	
Late	50	60.2
Early	33	39.8
Total known	83	100.0
Age at menopause (y)	Median age: 48 (42.45 – 53.56)	
Early	32	60.4
Late	21	39.6
Total known	53	100.0

Table 7 (continued): Overall distributions of clinicopathological and tumor features collected from the patients' medical files

Parameter	n	%
Family History of breast cancer		
No	53	66.3
Yes	27	33.7
Total known	80	100.0
Body Mass Index (BMI)		
Healthy	43	69.4
Obese	19	30.6
Total known	62	100.0
In situ component		
No	74	67.9
Yes	35	32.1
Total known	109	100.0
Histology		
Ductal	81	91.0
Not ductal	8	9.0
Total known	89	100.0
Histological grade		
Low/Intermediate	73	71.6
High	29	28.4
Total known	102	100.0
PgR Status		
Negative	19	17.9
Positive	87	82.0
Total known	106	100.0
HER2 Status		
Negative	16	44.4
Positive	20	55.6
Total known	36	100.0
Tumor size		
Small	46	42.6
Large	62	57.4
Total known	108	100.0
Lymph node involvement		
Low	81	75.7
High	26	24.3
Total known	107	100.0
Distant Metastases		
Absent	100	93.5
Present	7	6.5
Total known	107	100.0
Tumor Stage		
Early	73	68.9
Advanced	33	31.1
Total known	106	100.0

2.1. Allelic distributions stratified for patient and tumor characteristics

We then analyzed the allelic distributions throughout the characteristics collected from the patients to assess whether any of the alleles is overrepresented in any of the subsets and affects any of the features.

2.1.1. SNPs

We assessed the distributions of the variant alleles of each one of the SNPs studied in the same characteristics described in table 7. The results are shown in table 8.

Table 8: Distributions of patients' characteristics according to presence or absence of the variant allele of each of the studied SNPs

Parameter	rs4646		rs10046		rs700518		rs6493497	
	GA/AA n(%)	GG n(%)	AG/GG n(%)	AA n(%)	CT/TT n(%)	CC n(%)	GA/AA n(%)	GG n(%)
Age at diagnosis (y)								
>50	32 (66.7)	39 (63.9)	53 (66.3)	11 (37.9)	55 (68.8)	16 (55.2)	20 (76.9)	51 (63.8)
<50	16 (33.3)	22 (36.1)	27 (33.8)	18 (62.1)	25 (31.3)	13 (44.8)	6 (23.1)	29 (36.3)
P value^a	0.766		0.686		0.189		0.215	
Menopausal status								
Postmenopausal	26 (63.4)	35 (66.0)	44 (64.7)	17 (65.4)	46 (67.6)	15 (57.7)	18 (81.8)	26 (37.7)
Premenopausal	15 (36.6)	18 (34.0)	24 (35.3)	9 (34.6)	22 (32.4)	11 (42.3)	4 (18.2)	43 (62.3)
P value^a	0.792		0.951		0.366		0.090	
Age at menarche								
Late	12 (33.3)	21 (44.7)	35 (57.4)	15 (68.2)	37 (58.7)	13 (65.0)	11 (52.4)	38 (63.3)
Early	24 (66.7)	26 (55.3)	26 (42.6)	7 (31.8)	26 (41.3)	7 (35.0)	10 (47.6)	22 (36.7)
P value^a	0.295		0.375		0.618		0.377	
Age at menopause								
Early	15 (60.0)	17 (60.7)	26 (63.4)	6 (50.0)	26 (61.9)	6 (54.5)	10 (66.7)	22 (57.9)
Late	10 (40.0)	11 (39.3)	15 (36.6)	6 (50.0)	16 (38.1)	5 (45.5)	5 (33.3)	16 (42.1)
P value^a	0.958		0.403		0.657		0.556	
Family History of breast cancer								
No	25 (71.4)	28 (62.2)	38 (64.4)	15 (71.4)	37 (63.8)	16 (72.7)	12 (60.0)	38 (66.7)
Yes	10 (28.6)	17 (37.8)	21 (35.6)	6 (28.6)	21 (36.2)	6 (27.3)	8 (40.0)	19 (33.3)
P value^a	0.388		0.559		0.451		0.591	
Body Mass Index (BMI)								
Healthy	21 (80.8)	22 (61.1)	32 (69.6)	11 (68.8)	31 (68.9)	12 (70.6)	9 (60.0)	31 (70.5)
Obese	5 (19.2)	14 (38.9)	14 (30.4)	5 (31.3)	14 (31.1)	5 (29.4)	6 (40.0)	13 (29.5)
P value^a	0.098		0.951		0.897		0.454	
In situ component								
No	29 (60.4)	45 (73.8)	54 (67.5)	20 (69.0)	56 (70.0)	20 (69.0)	18 (69.2)	53 (74.6)
Yes	19 (39.6)	16 (26.2)	26 (32.5)	9 (31.0)	24 (30.0)	9 (31.0)	8 (30.8)	18 (25.4)
P value^a	0.138		0.885		0.885		0.779	
Histology								
Ductal	37 (88.1)	44 (93.6)	62 (91.2)	19 (90.5)	63 (91.3)	18 (90.0)	22 (100.0)	56 (87.5)
Not ductal	5 (11.9)	3 (6.4)	6 (8.8)	2 (9.5)	6 (8.7)	2 (10.0)	0 (0.0)	8 (12.5)
P value^a	0.363		0.922		0.857		0.082	

Table 8 (continued): Distributions of patients' characteristics according to presence or absence of the variant allele of each of the studied SNPs

Parameter	rs4646		rs10046		rs700518		rs6493497	
	GA/AA n(%)	GG n(%)	AG/GG n(%)	AA n(%)	CT/TT n(%)	CC n(%)	GA/AA n(%)	GG n(%)
Histological grade								
Low/Intermediate	33 (71.7)	40 (71.4)	54 (70.1)	19 (76.0)	54 (71.1)	19 (73.1)	19 (82.6)	53 (69.7)
High	13 (28.3)	16 (28.6)	23 (29.9)	6 (24.0)	22 (28.9)	7 (26.9)	4 (17.4)	23 (30.3)
<i>P</i> value ^a	0.972		0.572		0.843		0.225	
PgR Status								
Negative	6 (12.5)	13 (22.4)	15 (18.8)	4 (15.4)	16 (20.3)	3 (11.1)	5 (20.0)	14 (17.9)
Positive	42 (87.5)	45 (77.6)	65 (81.3)	22 (84.6)	63 (79.7)	24 (88.9)	20 (80.0)	64 (82.1)
<i>P</i> value ^a	0.185		0.698		0.285		0.818	
HER2 Status								
Negative	7 (41.2)	9 (47.4)	13 (46.4)	3 (37.5)	11 (40.7)	5 (55.6)	4 (40.0)	12 (50.0)
Positive	10 (58.8)	10 (52.6)	15 (53.6)	5 (62.5)	16 (59.3)	4 (44.4)	6 (60.0)	12 (50.0)
<i>P</i> value ^a	0.709		0.654		0.439		0.595	
Tumor size								
Small	20 (41.7)	26 (43.3)	31 (39.2)	15 (51.7)	33 (41.8)	13 (44.8)	12 (48.0)	33 (41.3)
Large	28 (58.3)	34 (56.7)	48 (60.8)	14 (48.3)	46 (58.2)	16 (55.2)	13 (52.0)	47 (58.8)
<i>P</i> value ^a	0.862		0.245		0.776		0.552	
Lymph node involvement								
Low	36 (76.6)	45 (75.0)	57 (73.1)	24 (82.8)	59 (75.6)	22 (75.9)	20 (80.0)	58 (73.4)
High	11 (23.4)	15 (25.0)	21 (26.9)	5 (17.2)	19 (24.4)	7 (24.1)	5 (20.0)	21 (26.6)
<i>P</i> value ^a	0.848		0.299		0.981		0.508	
Distant Metastases								
Absent	43 (93.5)	57 (93.4)	74 (94.9)	26 (89.7)	74 (94.9)	26 (89.7)	25 (96.2)	73 (93.6)
Present	3 (6.5)	4 (6.6)	4 (5.1)	3 (10.3)	4 (5.1)	3 (10.3)	1 (3.8)	5 (6.4)
<i>P</i> value ^a	0.274		0.429		0.429		0.637	
Tumor Stage								
Early	32 (71.1)	41 (67.2)	53 (68.8)	20 (69.0)	55 (71.4)	18 (62.1)	18 (69.2)	53 (68.8)
Advanced	13 (28.9)	20 (32.8)	24 (31.2)	9 (31.0)	22 (28.6)	11 (37.9)	8 (30.8)	24 (31.2)
<i>P</i> value ^a	0.668		0.989		0.354		0.970	

^a *p*-value calculated through a χ^2 test, in which variables are not independent if $p < 0.050$.

In relation to rs4646, the variant allele was not associated with any of the parameters studied, except for a suggestive non-significant association with body mass index, in which carriers of the variant allele (A) were more frequently healthy in weight (80.8%, $p=0.098$), compared to non-carriers (19.2%). In the case of HER2 status, we did not find any association, but results in the literature are inconsistent. One study reports that carriers of the variant allele have more probabilities of having HER2-positive tumors [29] and two other studies report that there is no association between the variant allele and HER2 status [96, 105]. So this needs further confirmation, and our study is not very reliable on that factor due to the reduced number of cases from which we had that information.

Concerning rs10046, the variant allele was not associated with any of the parameters studied as well. This is consistent with other studies that also did not find any association between this SNP and age and menopausal status at diagnosis, body mass index, PgR and HER2 status, tumor size, lymph node involvement, histological grade and tumor stage [29, 96, 102, 105]. However, another study reports an association between the variant allele of rs10046 with larger and more advanced tumors [92]. In that study they divided tumors as large or small using 50mm as a cutoff, unlike us that used 20mm as a cutoff, and that may explain this difference.

Additionally, rs700518 was also not associated with any of the studied parameters. Our results are consistent with another study in relation to age at menarche [106], because in that study no association is reported between the presence or absence of the variant allele of rs700518 and age at menarche.

Finally, concerning rs6493497, we did not find any significant associations but we did find a trend for carriers of the variant allele to be postmenopausal at diagnosis (81.8%, $p=0.090$). In addition all the carriers of the variant allele had invasive ductal carcinoma ($p=0.082$), with no other association or trend observed. We did not find in the literature any study that evaluated the relationship of this polymorphism with patient characteristics.

2.1.2. Tetranucleotide repeat polymorphism

We did the same analysis for the four most common alleles found in our population of the tetranucleotide repetition polymorphism, the one with seven repetitions and TCT deletion, eleven, seven and eight repetitions. The results are shown in table 9.

Table 9: Distribution of patients' characteristics according to the presence or absence of each of the studied (TTTA)_n alleles

Parameter	(TTTA) _{7-3bp}		(TTTA) ₁₁		(TTTA) ₇		(TTTA) ₈	
	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)
Age at diagnosis (y)								
>50	38 (64.4)	22 (66.7)	18 (58.1)	42 (68.9)	13 (68.4)	48 (64.9)	13 (68.4)	61 (65.6)
<50	21 (35.6)	11 (33.3)	13 (41.9)	19 (31.1)	6 (31.6)	26 (35.1)	6 (31.6)	32 (34.4)
P value^a	0.827		0.304		0.771		0.771	
Menopausal status								
Postmenopausal	29 (59.2)	22 (71.0)	18 (58.1)	33 (67.3)	13 (76.5)	38 (54.9)	10 (66.7)	41 (65.1)
Premenopausal	20 (40.8)	9 (29.0)	13 (41.9)	16 (32.7)	4 (23.5)	23 (45.1)	5 (33.3)	22 (34.9)
P value^a	0.285		0.400		0.277		0.908	
Age at menarche								
Late	25 (54.3)	15 (65.2)	13 (54.2)	27 (60.0)	10 (71.4)	35 (59.3)	11 (84.6)	34 (56.7)
Early	21 (45.7)	8 (34.8)	11 (45.8)	18 (40.0)	4 (28.6)	24 (40.7)	2 (15.4)	26 (43.3)
P value^a	0.389		0.640		0.402		0.060	
Age at menopause								
Early	16 (59.3)	10 (58.8)	5 (38.5)	21 (67.7)	9 (69.2)	19 (59.4)	5 (55.6)	23 (63.9)
Late	11 (40.7)	7 (41.2)	8 (61.5)	10 (32.3)	4 (30.8)	13 (40.6)	4 (44.4)	13 (36.1)
P value^a	0.977		0.071		0.537		0.645	
Family History of breast cancer								
No	27 (64.3)	17 (70.8)	17 (73.8)	27 (62.8)	9 (64.3)	34 (63.0)	8 (57.1)	35 (64.8)
Yes	15 (35.7)	7 (29.2)	6 (26.1)	16 (37.2)	5 (35.7)	20 (37.0)	6 (42.9)	19 (35.2)
P value^a	0.587		0.361		0.927		0.596	
Body Mass Index (BMI)								
Healthy	22 (68.8)	13 (65.0)	12 (66.7)	23 (67.6)	8 (61.5)	29 (74.4)	5 (83.3)	32 (69.6)
Obese	10 (31.3)	7 (35.0)	6 (33.3)	11 (32.4)	5 (38.5)	10 (25.6)	1 (16.7)	14 (30.4)
P value^a	0.779		0.943		0.377		0.484	
In situ component								
No	39 (66.1)	23 (69.7)	19 (61.3)	43 (70.5)	15 (78.9)	47 (63.5)	11 (57.9)	51 (68.9)
Yes	20 (33.9)	10 (30.3)	12 (38.7)	18 (29.5)	4 (21.1)	27 (36.5)	8 (42.1)	23 (31.1)
P value^a	0.724		0.374		0.203		0.363	
Histology								
Ductal	44 (91.7)	23 (92.0)	19 (86.4)	48 (94.1)	14 (93.3)	57 (89.1)	14 (87.5)	57 (90.5)
Not ductal	4 (8.3)	2 (8.0)	3 (13.6)	3 (5.9)	1 (6.7)	7 (10.9)	2 (12.5)	6 (9.5)
P value^a	0.961		0.286		0.622		0.725	

Table 9 (continued): Distribution of patients' characteristics according to the presence or absence of each of the studied (TTTA)_n alleles

Parameter	(TTTA) _{7-3bp}				(TTTA) ₇				(TTTA) ₈			
	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)	Carriers n(%)	Non-carriers n(%)
Histological grade												
Low/Intermediate	40 (70.2)	23 (76.7)	21 (72.4)	42 (72.4)	10 (62.5)	51 (70.8)	16 (88.9)	45 (64.3)				
High	17 (29.8)	7 (23.3)	8 (27.6)	16 (27.6)	6 (37.5)	21 (29.2)	2 (11.1)	25 (35.7)				
P value^a	0.520		1.000		0.513		0.044					
PgR Status												
Negative	13 (22.0)	3 (9.4)	3 (10.0)	13 (21.3)	3 (15.8)	15 (20.8)	1 (5.3)	17 (23.6)				
Positive	46 (78.0)	29 (90.6)	27 (90.0)	48 (78.7)	16 (84.2)	57 (79.2)	18 (94.7)	55 (76.4)				
P value^a	0.130		0.183		0.623		0.074					
HER2 Status												
Negative	10 (45.5)	5 (50.0)	5 (41.7)	10 (50.0)	3 (42.9)	10 (41.7)	1 (16.7)	12 (48.0)				
Positive	12 (54.5)	5 (50.0)	7 (58.3)	10 (50.0)	4 (57.1)	14 (58.3)	5 (83.3)	13 (52.0)				
P value^a	0.811		0.647		0.955		0.162					
Tumor size												
Small	24 (40.7)	16 (50.0)	17 (54.8)	23 (38.3)	9 (50.0)	29 (39.2)	6 (31.6)	32 (43.8)				
Large	35 (59.3)	16 (50.0)	14 (45.2)	37 (61.7)	9 (50.0)	45 (60.8)	13 (68.4)	41 (56.2)				
P value^a	0.392		0.133		0.403		0.334					
Lymph node involvement												
Low	43 (74.1)	28 (87.5)	26 (86.7)	45 (75.0)	16 (88.9)	52 (71.2)	14 (73.7)	54 (75.0)				
High	15 (25.9)	4 (12.5)	4 (13.3)	15 (25.0)	2 (11.1)	21 (28.8)	5 (26.3)	18 (25.0)				
P value^a	0.137		0.201		0.123		0.907					
Distant Metastases												
Absent	55 (96.5)	30 (90.9)	29 (93.5)	56 (94.9)	17 (89.5)	69 (95.8)	17 (100.0)	69 (93.2)				
Present	2 (3.5)	3 (9.1)	2 (6.5)	3 (5.1)	2 (10.5)	3 (4.2)	0 (0.0)	5 (6.8)				
P value^a	0.302		0.573		0.279		0.270					
Tumor Stage												
Early	41 (73.2)	23 (69.7)	22 (73.3)	42 (71.2)	13 (68.4)	50 (70.4)	12 (70.6)	51 (69.9)				
Advanced	15 (26.7)	10 (30.3)	8 (26.7)	17 (28.8)	6 (31.6)	21 (29.6)	5 (29.4)	22 (30.1)				
P value^a	0.721		0.831		0.866		0.953					

^a p-value calculated through a X² test, in which variables are not independent if p<0.050.

For the seven repetitions and TCT deletion allele, no association was observed with any of the studied parameters. In terms of menopausal status at diagnosis, our results are consistent with two other studies [31, 32], which report a similar frequency of premenopausal and postmenopausal patients within carriers of this allele. However, another study reports that this allele is significantly overrepresented in premenopausal patients as well as in women diagnosed before 50 [98]. Maybe this difference can be due to the fact that in this study were used pieces of paraffin-embedded breast tissue instead of blood, which can affect the quality of the DNA isolated and the results of the genotyping.

On the other hand, concerning the eleven repetitions allele, we only observed a statistical trend for carriers to have menopause at a later age (61.5%, $p=0.071$). However, the total number of carriers for which we had that information was only 13, so we cannot make a sure assumption from that. A study [107] reports an association between carriers of this allele and late onset of menarche. In our study we have slightly more patients with late menarche among the carriers, but this association is not significant ($p=0.640$). This difference may be due to the small number of carriers from which we had that information which was only 24. However, the same study reports no association between the presence of this allele and body mass index, which is consistent with our results.

Concerning the seven repetition polymorphism, according to our findings, the presence of this allele is not associated with any of the studied parameters. This is in accordance with another study that reports no significant difference in distribution between carriers and non-carriers in terms of menopausal status at diagnosis, histology of the tumor, presence of lymph node involvement, tumor size, histological grade, hormone receptor status and adjuvant chemotherapy [100]. However, as for the eleven repetition allele, in another study it was associated with age at menarche, and this time with early menarche [107]. On the contrary, in our study most of the carriers of this allele (71.4%, $p=0.402$) had late menarche. Again, the number of carriers of the allele in our population is very small, only 14, but there may be another additional unknown reason for this difference. Also the same study reported that this allele is not associated with body mass index, which is consistent with the results observed in our study.

Finally, regarding the eight repetition allele, in our study its presence has a statistical trend to be associated with late menarche (84.6% of the carriers, $p=0.060$). Additionally, it is significantly over-represented in low to intermediate histological

grade tumors (88.9%, $p=0.044$). And also the presence of this allele has a statistical trend to be over-represented in PgR-positive tumors (94.7%, $p=0.074$).

3. Survival Analysis

The main aim of our study was to assess whether any of the studied polymorphisms has any influence in the response to aromatase inhibitor treatment. For that, we performed Kaplan-Meier survival analyses, which are used to measure the time to onset of any endpoint, which can be death or progression, as in the context of our study. So we studied the overall survival of patients, or time until death, and progression free survival, or time until progression. The total follow-up time for all of these analyses was 120 months.

3.1. *Global survival analysis*

First we assessed the overall and progression free survival in the global population, without any stratification. As shown in figure 10 the overall survival is 88.1%, which means that at the end of the 120 months of follow-up 88.1% of the patients were still alive, with a mean time until death of 113.6 months (± 1.819 months, 95% CI). On the other hand, in terms of progression free survival, we observed that at the end of the follow-up time, 64.2% of the patients did not have progressed or suffered recurrences of their breast cancers, with a mean time to progression of 94.4 months (± 3.755 months, 95% CI).

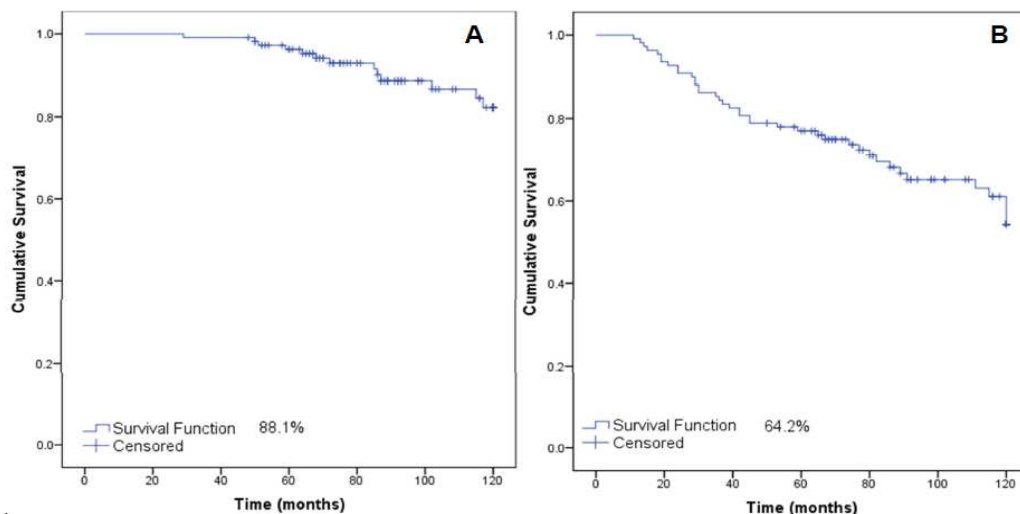


Figure 10: Overall and progression free survival obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival.

After that, we stratified the survival analysis for clinical features and treatments applied, to assess whether any subset of the population had any significant difference in terms of overall and progression free survival. The results are shown in table 10.

In this analysis we included all the features analyzed for the allelic distributions, as well as treatments applied. These included neoadjuvant treatment, or treatment applied before surgery, adjuvant chemotherapy and radiotherapy, which can affect the prognosis. Also, we added the use of oral contraceptives and breast feeding to the analysis as well, because they may also affect the survival.

So, from the analysis of table 10, we observed that neither overall survival nor progression free survival was affected by any of the clinical features. However, overall, patients with diagnosis after 50 years of age, postmenopausal, with early menarche and late menopause, no use of oral contraceptives, no family history of breast cancer and healthy body mass index, had a higher overall and progression free survival. Concerning breast feeding, however, those patients who breast fed their children had a lower overall, but higher progression free survival.

On the other hand, concerning tumor features we observed that patients with lower histological grade tumors had a significantly higher overall survival (93.2% over 75.9%, $p=0.021$), but no significant differences in progression free survival (67.1% for lower histological grade over 58.6%, $p=0.534$). Additionally, patients who had tumors with distant metastases at diagnosis had also a significantly ($p<0.001$) lower overall (57.1% over 90.0%) and progression free (14.3% over 68.0%) survival. For the other tumor features we did not observe any significant difference, however, presence of *in situ* component associated with the primary tumor, histology of ductal carcinoma, PgR positivity, smaller tumors, low lymph node involvement and earlier tumor stage, were all not significantly associated with a higher overall and progression free survival, which is consistent with the presence of less aggressive tumors. Regarding HER2 status, however, positivity was not significantly associated with a better overall survival but worse progression free survival. However, this feature is biased, because it was only assessed for patients who needed to be treated with Trastuzumab (Herceptin®), after failure of hormone therapy, so these patients are more likely to have more aggressive tumors. These results are consistent with other studies that report that positivity of PgR [105] and HER2 [108] are associated with a better response to treatment.

Table 10: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied

Covariables	Overall Survival				Progression Free Survival			
	n	Events	OS (%)	<i>p</i> ^a	n	Events	PFS (%)	<i>p</i> ^a
Age at Diagnosis (y)								
>50	70	6	91.4	0.201	70	21	70.0	0.235
<50	39	7	82.1		39	18	53.8	
Menopausal Status								
Post-M ^b	61	6	90.2	0.372	61	17	72.1	0.342
Pre-M ^b	33	6	81.8		33	15	54.5	
Age at Menarche								
Late	50	7	86.0	0.980	50	20	60.0	0.874
Early	33	4	87.9		33	11	66.7	
Age at Menopause								
Early	32	5	84.4	0.633	32	7	78.1	0.166
Late	21	2	90.5		21	8	61.9	
Breast Feeding								
No	9	1	88.9	0.891	9	4	55.6	0.731
Yes	24	3	87.5		24	8	66.7	
Contraceptives								
No	21	3	85.7	0.553	21	8	61.9	0.261
Yes	29	5	82.8		29	14	51.7	
Family history of breast cancer								
No	53	7	86.8	0.855	53	19	64.2	0.319
Yes	27	4	85.2		27	13	51.9	
Body Mass Index								
Healthy	43	5	88.4	0.380	43	14	67.4	0.252
Obese	19	4	78.9		19	10	47.4	
In situ component								
No	74	11	85.1	0.349	74	29	60.8	0.864
Yes	35	2	94.3		35	10	71.4	
Histology								
Ductal	81	10	87.7	0.821	81	30	63.0	0.231
Not ductal	8	1	87.5		8	2	75.0	
Histological grade								
Low/Inter ^c	73	5	93.2	0.021	73	24	67.1	0.534
High	29	7	75.9		29	12	58.6	
PgR Status								
Negative	19	4	78.9	0.199	19	6	68.4	0.683
Positive	87	8	90.8		87	31	64.4	
HER2 Status								
Negative	16	5	68.8	0.940	16	10	37.5	0.387
Positive	20	5	75.0		20	15	25.0	
Tumor size								
Small	46	3	93.5	0.350	46	13	71.7	0.724
Large	62	9	85.5		62	25	59.7	
Lymph node involvement								
Low	81	8	90.1	0.574	81	26	67.9	0.485
High	26	4	84.6		26	12	53.8	
Distant Metastases								
Absent	100	10	90.0	<0.001	100	32	68.0	<0.001
Present	7	3	57.1		7	6	14.3	
Tumor Stage								
Early	73	6	91.8	0.102	73	23	68.5	0.449
Advanced	33	7	78.8		33	15	54.5	

Table 10 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied

Covariables	Overall Survival				Progression Free Survival			
	n	Events	OS (%)	<i>p</i> ^a	n	Events	PFS (%)	<i>p</i> ^a
Neoadjuvant Treatment								
No	98	11	88.8	0.448	98	34	65.3	0.994
Yes	10	2	80.0		10	4	60.0	
Adjuvant Chemotherapy								
No	31	3	90.3	0.903	31	8	74.2	0.581
Yes	78	10	87.2		78	31	60.3	
Adjuvant Radiotherapy								
No	17	4	76.5	0.034	17	8	52.9	0.042
Yes	92	9	90.2		92	31	66.3	

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if $p < 0.050$, the results are significantly different; ^b M – menopausal; ^c Inter. – Intermediate.

Finally, concerning treatments applied, we observed that patients who were administered adjuvant radiotherapy had a significantly better overall survival (90.2% over 76.5%, $p=0.034$) and progression free survival (66.3% over 52.9%, $p=0.042$). On the other hand, interestingly patients who were administered neoadjuvant treatment and adjuvant chemotherapy had a not significantly lower overall survival and higher progression free survival.

3.2. Survival analysis for the SNPs

We did the same analysis comparing the presence and absence of the variant allele of the SNPs, to assess whether the presence of these alleles affected the survival any further.

3.2.1. *rs4646*

Concerning the overall and progression free survival not stratified, the results are shown in figure 11. We did not observe any significant difference in the presence or absence of the variant allele of *rs4646* in terms of overall (86.9% for G and 89.6% for A, $p=0.806$, figure 11-A) or progression free (63.9% for G and 64.6% for A, $p=0.790$, figure 11-B) survival. This is consistent with another study that reports that this SNP is not associated with overall survival [29]. However, another study reports a 3-fold increase in progression free survival for carriers of the variant allele of this SNP [105]. This difference may be due to the fact that in that study has a reported follow up time of 40 months, as opposed to our 120 months, and additionally their population is smaller than ours, including only 67 patients, all of them with progression after first line treatment. Yet another study reports a worse progression free survival for carriers of the variant allele of this SNP [96], but the patients included

in that study were treated with letrozole before surgery, as a neoadjuvant therapy, which can affect the survival results, and make them not so comparable with ours.

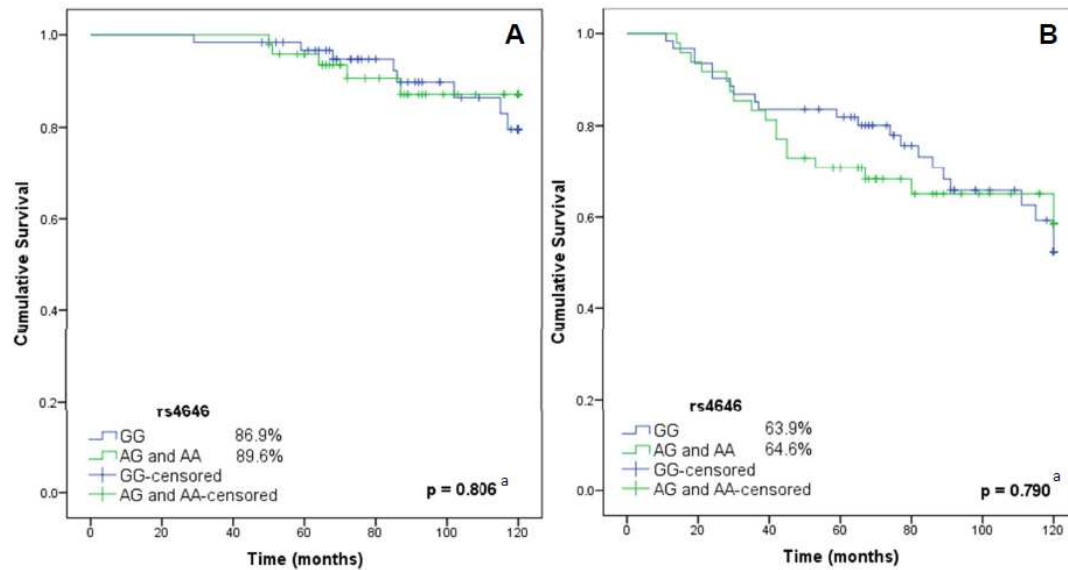


Figure 11: Overall and progression free survival for the presence or absence of the variant allele of rs4646 obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

Subsequently, we stratified the analysis for clinical and tumor features to verify if the association was still not significant or any of the subsets had a significant difference in survival. The results are shown in table 11.

Concerning overall survival, we did not find any significant difference in the different subsets of patients for the presence or absence of the variant allele of rs4646. However, in terms of progression free survival, we observed that carriers of the variant allele of rs4646 who never used oral contraceptives had a significantly lower progression free survival, compared to non-carriers (20.0% over 75.0%, $p = 0.006$). Additionally, progression free survival was also lower in carriers of the variant allele of rs4646 with presence of an *in situ* component associated with the primary tumor (52.6% over 93.8%, $p = 0.011$) and higher for carriers of this allele with not ductal carcinomas (100.0% over 33.3%, $p = 0.017$). For the other parameters there was no significant difference in progression free survival, except for a statistical trend for patients who never breast fed to have a worse progression free survival (33.3% over 100.0%, $p = 0.093$). This association was not significant probably due to the reduced number of patients from which we obtained this information, which was only 19.

Table 11: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs4646

Covariables	Overall Survival						Progression Free Survival							
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele				
	n	Events	OS (%)	n	Events	OS (%)	p ^a	n	Events	PFS (%)	n	Events	PFS (%)	p ^a
Age at Diagnosis (y)														
>50	31	2	93.5	39	4	89.7	0.768	31	9	71.0	39	12	69.2	0.789
<50	17	3	82.4	22	4	81.9	0.934	17	8	52.9	22	10	54.5	0.949
Menopausal Status														
Post-M ^b	26	2	92.3	35	4	88.6	0.793	26	7	73.1	35	10	71.4	0.893
Pre-M ^b	15	3	80.0	18	3	83.3	0.738	15	8	46.7	18	7	61.1	0.424
Age at Menarche														
Late	24	3	87.5	26	4	84.6	0.901	24	11	54.2	26	9	65.4	0.289
Early	12	2	83.3	21	2	90.5	0.376	12	4	66.7	21	7	66.7	0.572
Age at Menopause														
Early	15	2	86.7	17	3	82.4	0.935	15	3	80.0	17	4	76.5	0.946
Late	10	1	90.0	11	1	90.9	0.294	10	5	50.0	11	3	72.7	0.139
Breast Feeding														
No	6	1	83.3	3	0	100.0	0.617	6	4	33.3	3	0	100.0	0.093
Yes	7	1	85.7	17	2	88.2	0.793	7	2	71.4	17	6	64.7	0.857
Contraceptives														
No	5	1	80.0	16	2	87.5	0.619	5	4	20.0	16	4	75.0	0.006
Yes	13	3	76.9	16	2	87.5	0.503	13	8	38.5	16	6	62.5	0.242
Family history of breast cancer														
No	25	4	84.0	28	3	89.3	0.402	25	10	60.0	28	9	67.9	0.313
Yes	10	1	90.0	17	3	82.4	0.694	10	4	60.0	17	9	47.1	0.476
Body Mass Index														
Healthy	21	2	90.5	22	3	86.4	0.731	21	6	71.4	22	8	63.6	0.574
Obese	5	0	100.0	14	4	71.4	0.383	5	1	80.0	14	9	35.7	0.276
In situ component														
No	29	3	89.7	45	8	82.2	0.537	29	8	72.4	45	21	53.3	0.214
Yes	19	2	89.5	16	0	100.0	0.251	19	9	52.6	16	1	93.8	0.011
Histology														
Ductal	37	4	89.2	44	6	86.4	0.903	37	15	59.5	44	15	65.9	0.310
Not ductal	5	0	100.0	3	1	66.7	0.114	5	0	100.0	3	2	33.3	0.017
Histological Grade														
Low/Inter ^c	33	2	93.9	40	3	92.5	0.874	33	10	69.7	40	14	65.0	0.768
High	13	3	76.9	16	4	75.0	0.941	13	7	46.2	16	5	68.8	0.161

Table 11 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs4646

Covariables	Overall Survival					Progression Free Survival				
	Variant Allele			Wild-Type Allele		p ^a	Variant Allele			p ^a
	n	Events	OS (%)	n	Events	OS (%)	n	Events	PFS (%)	PFS (%)
PgR Status										
Negative	6	1	83.3	13	3	76.9	6	1	83.3	61.5
Positive	42	4	90.5	45	4	91.1	42	16	61.9	66.7
										0.496
										0.548
HER2 Status										
Negative	7	2	71.4	9	3	66.7	7	4	42.9	33.3
Positive	10	3	70.0	10	2	80.0	10	7	30.0	20.0
										0.867
										0.832
Tumor size										
Small	20	1	95.0	26	2	92.3	20	4	80.0	65.4
Large	28	4	85.7	34	5	85.3	28	13	53.6	64.7
										0.613
										0.356
Lymph node involvement										
No	36	4	88.9	45	4	91.1	36	12	66.7	68.9
Yes	11	1	90.9	15	3	80.0	11	5	54.5	53.3
										0.557
										0.866
Distant Metastases										
Absent	43	4	90.7	57	6	89.5	43	13	69.8	66.7
Present	3	1	66.7	4	2	50.0	3	3	0.0	25.0
										0.948
										0.355
Tumor Stage										
Early	32	3	90.6	41	3	92.7	32	11	65.6	70.7
Advanced	13	2	84.6	20	5	75.0	13	5	61.5	50.0
										0.342
										0.564
Neoadjuvant Treatment										
No	41	4	90.2	57	7	87.7	41	13	68.3	63.2
Yes	6	1	83.3	4	1	75.0	6	3	50.0	75.0
										0.810
										0.355
Adjuvant Chemotherapy										
No	13	1	92.3	18	2	88.9	13	4	69.2	77.8
Yes	35	4	88.6	43	6	86.0	35	13	62.9	58.1
										0.402
										0.849
Adjuvant Radiotherapy										
No	4	0	100.0	13	4	69.2	4	1	75.0	46.2
Yes	44	5	88.6	48	4	91.7	44	16	63.6	68.8
										0.477
										0.419

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if p<0.050, the results are significantly different; ^b M – menopausal; ^c Inter. – Intermediate.

Therefore, we observed a slight tendency for carriers of the variant allele of rs4646 to have a worse response to the treatment. A study reports that the variant allele of this SNP is associated with increased estrogen levels [101] and this might be the reason for the tendency of carriers to have a worse response to the treatment, because increased estrogen levels may be associated with increased aromatase levels and thus more substrate for the aromatase inhibitors, rendering them less effective.

3.2.2. *rs10046*

The results for the not stratified survival analysis for rs10046 are shown in figure 12. We observed that there was no significant difference in terms of overall (89.7% for the wild-type allele and 87.5% for the variant allele, $p=0.926$, Figure 12-A) and progression free (62.1% for the wild-type allele and 65.0% for the variant allele, $p=0.728$, Figure 12-B). This is consistent with another study that also reports no significant difference in progression free survival between carriers and non-carriers of the variant allele of rs10046 [105]. Also, another study reports no effect on the presence or absence of the variant allele of rs10046 on survival [95]. However, another study reports a statistical trend for the variant allele of this SNP to be associated with a better progression free survival but not overall survival [29]. This difference may be due to the short follow up time (50 months), which may not be enough to draw sure conclusions.

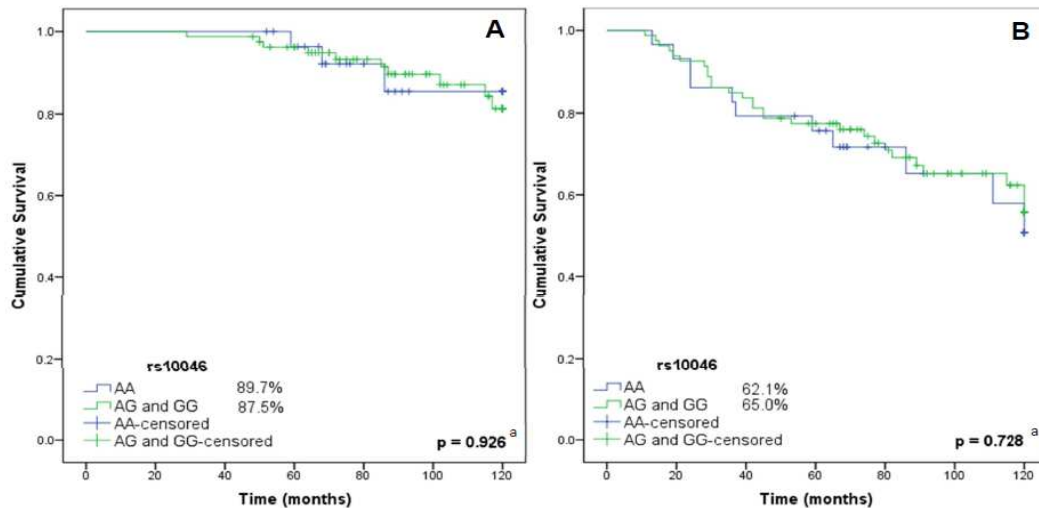


Figure 12: Overall and progression free survival for the presence or absence of the variant allele of rs10046 obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

We stratified the analysis for the same parameters analyzed for rs4646, and the results are shown in table 12.

Table 12: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs10046

Covariables		Overall Survival						Progression Free Survival							
		Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele				
		n	Events	OS (%)	n	Events	OS (%)	n	Events	PFS (%)	n	Events	PFS (%)		
Age at Diagnosis (y)															
>50	52	5	90.4	18	1	94.4	0.767								
<50	28	5	82.1	11	2	81.8	0.954								
Menopausal Status															
Post-M ^b	44	5	88.6	17	1	94.1	0.632								
Pre-M ^b	24	5	79.2	9	1	88.9	0.562								
Age at Menarche															
Late	35	6	82.9	15	1	93.3	0.477								
Early	26	3	88.5	7	1	85.7	0.893								
Age at Menopause															
Early	26	4	84.6	6	1	83.3	0.834								
Late	15	2	86.7	6	0	100.0	0.392								
Breast Feeding															
No	8	1	87.5	1	0	100.0	--								
Yes	15	3	80.0	9	0	100.0	0.199								
Contraceptives															
No	11	3	72.7	10	0	100.0	0.229								
Yes	24	4	83.3	5	1	80.0	0.963								
Family history of breast cancer															
No	38	6	84.2	15	1	93.3	0.460								
Yes	21	3	85.7	6	1	83.3	0.926								
Body Mass Index															
Healthy	22	3	90.6	11	2	81.8	0.355								
Obese	14	4	71.4	5	0	100.0	0.226								
In situ component															
No	54	8	85.2	20	3	85.0	0.997								
Yes	26	2	82.3	9	0	100.0	0.512								
Histology															
Ductal	62	8	87.1	19	2	89.5	0.756								
Not ductal	6	0	100.0	2	1	50.0	0.014								
Histological Grade															
Low/Inter ^c	54	4	92.6	19	1	94.7	0.947								
High	23	6	73.9	6	1	83.3	0.807								

Table 12 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs10046

Covariables	Overall Survival						Progression Free Survival					
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele		
	n	Events	OS (%)	n	Events	OS (%)	n	Events	PFS (%)	n	Events	PFS (%)
Pgr Status												
Negative	15	3	80.0	4	1	75.0	15	4	73.3	4	2	50.0
Positive	65	7	89.2	22	1	95.5	65	24	63.1	22	7	68.2
						0.446						0.733
HER2 Status												
Negative	13	4	69.2	3	1	66.7	13	7	46.2	3	3	0.0
Positive	15	4	73.3	5	1	80.0	15	11	26.7	5	4	20.0
						0.691						0.863
Tumor size												
Small	31	2	93.5	15	1	93.3	31	7	77.4	15	6	60.0
Large	48	7	85.4	14	2	85.7	48	20	58.3	14	5	64.3
						0.793						0.285
Lymph node involvement												
No	57	6	89.5	24	2	91.7	57	17	70.2	24	9	62.5
Yes	21	3	85.7	5	1	80.0	21	10	52.4	5	2	60.0
						0.710						0.815
Distant Metastases												
Absent	74	8	89.2	26	2	92.3	74	23	68.9	26	9	65.4
Present	4	2	50.0	3	1	66.7	4	4	0.0	3	2	33.3
						0.619						0.174
Tumor Stage												
Early	53	5	90.6	20	1	95.0	53	16	69.8	20	7	65.0
Advanced	24	5	79.2	9	2	77.8	24	11	54.2	9	4	55.6
						0.862						0.943
Neoadjuvant Treatment												
No	73	9	87.7	25	2	92.0	73	24	67.1	25	10	60.0
Yes	6	1	83.3	4	1	75.0	6	3	50.0	4	1	75.0
						0.886						0.345
Adjuvant Chemotherapy												
No	21	2	90.5	10	1	90.0	21	6	71.4	10	2	80.0
Yes	59	8	86.4	19	2	89.5	59	22	62.7	19	9	52.6
						0.873						0.426
Adjuvant Radiotherapy												
No	9	2	77.8	8	2	75.0	9	4	55.6	8	4	50.0
Yes	71	8	88.7	21	1	95.2	71	24	66.2	21	7	66.7
						0.436						0.905

^aOverall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if $p < 0.050$, the results are significantly different; ^bM – menopausal; ^cInter. – Intermediate.

None of the clinical features was significantly affected by the presence or absence of the variant allele of rs10046 in terms of overall survival. However, carriers of this allele with not ductal carcinoma had a significantly higher overall survival (100.0% over 50.0%, $p=0.014$), with none of the other tumor characteristics and treatments administered affecting the survival for carriers and non-carriers of this allele.

On the other hand, in terms of progression free survival, we observed that carriers of the variant allele of rs10046 with a low body mass index had a significantly higher progression free survival than non-carriers (75.0% as opposed to 45.5%, $p=0.044$). Again, none of the other clinical, tumor and treatment features was significantly associated with progression free survival. A study reports that the progression free survival in premenopausal patients carrying the variant allele of rs10046 was better [29] however in that study the follow-up time was shorter than ours, which may be the reason for these differences.

Hence, we observed a better treatment response for carriers of the variant allele of this polymorphism. A study reports that the G allele of rs10046 is associated with lower estrogen levels [45] so if estrogen levels are already low, it may mean that the aromatase enzyme is less active or there are lower levels of aromatase, and consequently less target for the aromatase inhibitors, making them more effective.

3.2.3. rs700518

The overall and progression free survival for carriers and non-carriers of the variant allele of rs700518 is shown in figure 13. We did not observe any significant difference in overall and progression free survival. However, from the graphs in figure 13, we can see a non-significant trend for carriers of the variant allele of rs700518 to have a slightly better overall (90.0% over 82.8%, $p=0.302$) and progression free (67.5% over 55.2%, $p=0.332$) survival than non-carriers.

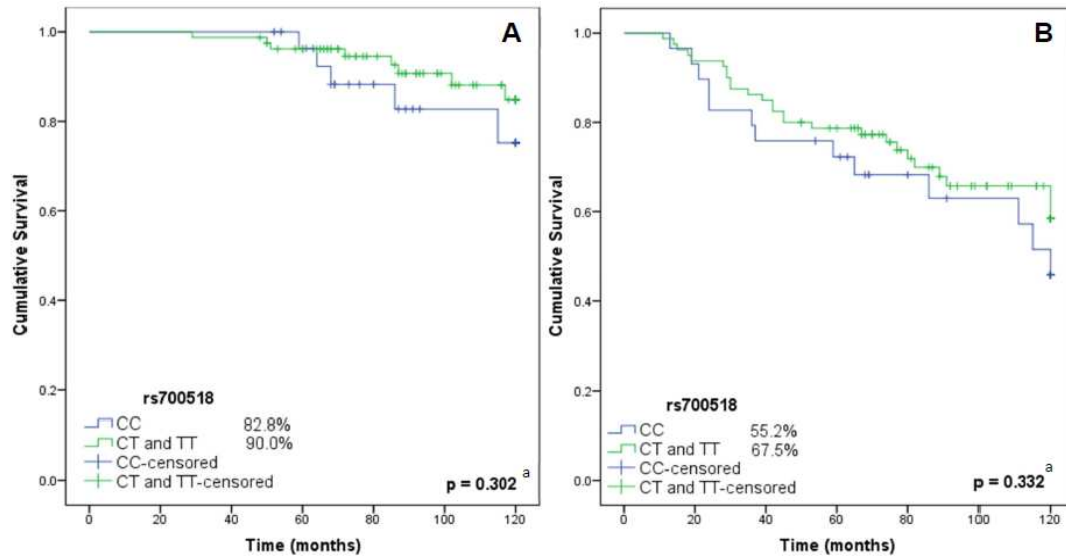


Figure 13: Overall and progression free survival for the presence or absence of the variant allele of rs700518 obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

Again, we stratified the analysis for clinical, tumor and treatment features, for a possible stronger association in any of the subsets. The results are shown in table 13.

In terms of overall survival we observed that, as for rs10046, only carriers of the variant allele with not ductal carcinoma had a significantly higher overall survival than non-carriers (100.0%, compared to 50.0%, $p=0.014$). Also, we observed a trend for carriers of the allele with a low body mass index to have a better overall survival than non-carriers (93.5% as opposed to 75.0, $p=0.055$).

On the other hand, in terms of progression free survival, carriers of the allele with a low body mass index had a significantly better progression free survival than non-carriers (77.4% over 41.7%, $p=0.009$). Also, carriers with PgR-positive tumors had a significantly better progression free survival, as well (75.0% compared to 33.3, $p=0.048$), with the same being observed for carriers of the allele with HER2 positive tumors (31.3% over 0.0%, $p=0.011$).

Table 13: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs700518

Covariables	Overall Survival						Progression Free Survival							
	Variant Allele			Wild-Type Allele			p ^a	Variant Allele			Wild-Type Allele			p ^a
	n	Events	OS (%)	n	Events	OS (%)		n	Events	PFS (%)	n	Events	PFS (%)	
Age at Diagnosis (y)														
>50	54	4	92.6	16	2	87.5	0.547	54	15	72.2	16	6	62.5	0.583
<50	26	4	84.6	13	3	76.9	0.547	26	11	57.7	13	7	46.2	0.531
Menopausal Status														
Post-M ^b	46	4	91.3	15	2	86.7	0.664	46	12	73.9	15	5	66.7	0.638
Pre-M ^b	22	4	81.8	11	2	81.8	0.987	22	10	54.5	11	5	54.5	0.979
Age at Menarche														
Late	37	5	86.5	13	2	84.6	0.799	37	16	56.8	13	4	69.2	0.450
Early	26	3	88.5	7	1	85.7	0.893	26	7	73.1	7	4	42.9	0.295
Age at Menopause														
Early	26	4	84.6	6	1	83.3	0.834	26	5	80.8	6	2	66.7	0.474
Late	16	1	93.8	5	1	80.0	0.690	16	6	62.5	5	2	60.0	0.672
Breast Feeding														
No	8	1	87.5	1	0	100.0	--	8	4	50.0	1	0	100.0	0.510
Yes	13	1	92.3	11	2	81.8	0.598	13	3	76.9	11	5	54.5	0.363
Contraceptives														
No	13	3	76.9	8	0	100.0	0.229	13	6	53.8	8	2	75.0	0.468
Yes	22	3	86.4	7	2	71.4	0.511	22	10	54.5	7	4	42.9	0.556
Family history of breast cancer														
No	37	4	89.2	16	3	81.3	0.492	37	12	67.6	16	7	56.3	0.452
Yes	21	3	85.7	6	1	83.3	0.926	21	9	57.1	6	4	33.3	0.386
Body Mass Index														
Healthy	31	2	93.5	12	3	75.0	0.055	31	7	77.4	12	7	41.7	0.009
Obese	14	3	78.6	5	1	80.0	0.651	14	7	50.0	5	3	40.0	0.776
In situ component														
No	54	7	87.0	20	4	80.0	0.637	54	18	66.7	20	11	45.0	0.327
Yes	26	1	96.2	9	1	88.9	0.166	26	8	69.2	9	2	77.8	0.754
Histology														
Ductal	63	7	88.9	18	3	83.3	0.820	63	22	65.1	18	8	55.6	0.756
Not ductal	6	0	100.0	2	1	50.0	0.014	6	1	83.3	2	1	50.0	0.307
Histological Grade														
Low/Inter ^c	54	3	94.4	19	2	89.5	0.415	54	15	72.2	19	9	52.6	0.162
High	22	5	77.3	7	2	71.4	0.716	22	9	59.1	7	3	57.1	0.796

Table 13 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs700518

Covariables	Overall Survival						Progression Free Survival							
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele				
	n	Events	OS (%)	n	Events	OS (%)	p ^a	n	Events	PFS (%)	n	Events	PFS (%)	p ^a
PgR Status														
Negative	16	3	81.3	3	1	66.7	0.406	16	4	75.0	3	2	33.3	0.048
Positive	63	5	92.1	24	3	87.5	0.571	63	21	66.7	24	10	58.3	0.663
HER2 Status														
Negative	11	3	72.7	5	2	60.0	0.884	11	5	54.5	5	5	0.0	0.424
Positive	16	3	81.3	4	2	50.0	0.309	16	11	31.3	4	4	0.0	0.011
Tumor size														
Small	33	2	93.9	13	1	92.3	0.923	33	8	75.8	13	5	61.5	0.486
Large	46	5	89.1	16	4	75.0	0.135	46	17	63.0	16	8	50.0	0.399
Lymph node involvement														
No	59	5	91.5	22	3	86.4	0.421	59	17	71.2	22	9	59.1	0.291
Yes	19	2	89.5	7	2	71.4	0.371	19	8	57.9	7	4	42.9	0.840
Distant Metastases														
Absent	74	6	91.9	26	4	84.6	0.293	74	21	71.6	26	11	57.7	0.271
Present	4	2	50.0	3	1	66.7	0.619	4	4	0.0	3	2	33.3	0.174
Tumor Stage														
Early	55	4	92.7	18	2	88.9	0.536	55	16	70.9	18	7	61.1	0.409
Advanced	22	4	81.8	11	3	72.7	0.638	22	9	59.1	11	6	45.5	0.754
Neoadjuvant Treatment														
No	73	7	90.4	25	4	84.0	0.298	73	22	69.9	25	12	52.0	0.114
Yes	6	1	83.3	4	1	75.0	0.886	6	3	50.0	4	1	75.0	0.355
Adjuvant Chemotherapy														
No	22	2	90.9	9	1	88.9	0.837	22	6	72.7	9	2	77.8	0.699
Yes	58	6	89.7	20	4	80.0	0.279	58	20	65.5	20	11	45.0	0.184
Adjuvant Radiotherapy														
No	9	2	77.8	8	2	75.0	0.819	9	4	55.6	8	4	50.0	0.578
Yes	71	6	91.5	21	3	85.7	0.502	71	22	69.0	21	9	57.1	0.531

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if p<0.050, the results are significantly different; ^b M – menopausal; ^c Inter. – Intermediate.

3.2.4. rs6493497

The results from the overall and progression free survival analysis for the presence or absence of the variant allele of rs6493497 are shown in figure 14.

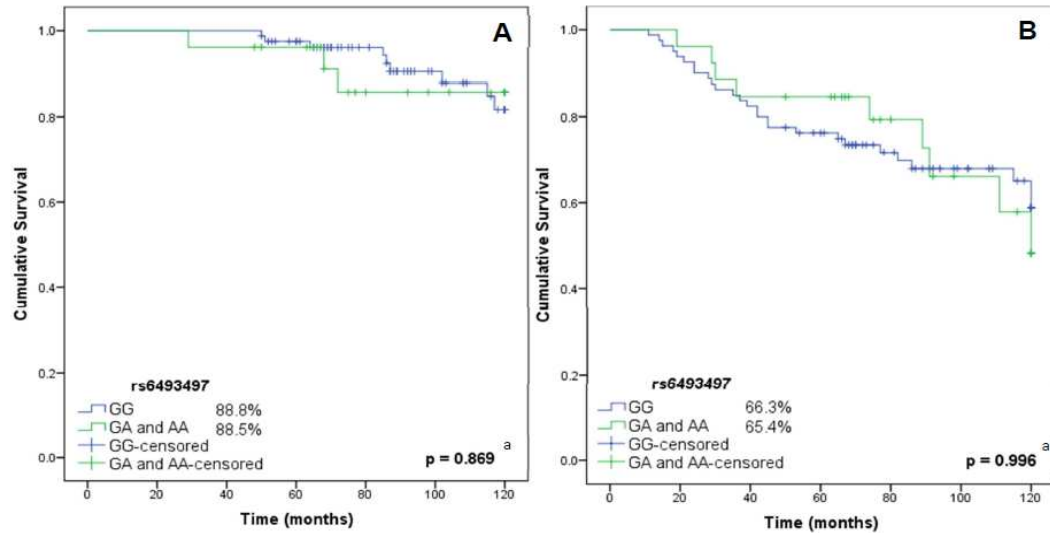


Figure 14: Overall and progression free survival for the presence or absence of the variant allele of rs6493497 obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

We observed that there is no significant difference between the presence and absence of the variant allele in terms of overall (88.8% in the absence and 88.5% in the presence of the allele, $p=0.869$) and progression free (66.3% for G and 65.4% for A, $p=0.996$). We then sought for significant differences if the analysis was stratified for the clinical, tumor and treatment features analyzed. The results are shown in table 14.

Table 14: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs6493497

Covariables	Overall Survival						Progression Free Survival								
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele					
	n	Events	OS (%)	n	Events	OS (%)	p ^a	n	Events	PFS (%)	n	Events	PFS (%)	p ^a	
Age at Diagnosis (y)	>50	20	1	95.0	50	5	90.0	0.516	20	5	75.0	50	16	68.0	0.529
	<50	6	2	66.7	30	4	86.7	0.175	6	4	33.3	30	11	63.3	0.186
Menopausal Status	Post-M ^b	18	2	88.9	43	4	90.7	0.725	18	4	77.8	43	13	69.8	0.542
	Pre-M ^b	4	1	75.0	26	4	84.6	0.650	4	3	25.0	26	9	65.4	0.156
Age at Menarche	Late	11	2	81.8	38	5	86.8	0.573	11	4	63.6	38	15	60.5	0.772
	Early	10	0	100.0	22	3	86.4	0.244	10	3	70.0	22	7	68.2	0.776
Age at Menopause	Early	10	3	70.0	22	2	90.9	0.105	10	3	70.0	22	4	81.8	0.390
	Late	5	0	100.0	16	2	87.5	0.422	5	0	100.0	16	8	50.0	0.074
Breast Feeding	No	2	0	100.0	6	1	83.3	--	2	0	100.0	6	3	50.0	0.267
	Yes	6	1	83.3	17	2	88.2	0.428	6	1	83.3	17	6	64.7	0.592
Contraceptives	No	4	0	700.0	17	3	82.4	0.446	4	0	100.0	17	8	52.9	0.129
	Yes	9	2	77.8	17	2	88.2	0.385	9	3	66.7	17	8	52.9	0.676
Family history of breast cancer	No	12	1	91.7	38	5	86.8	0.621	12	2	83.3	38	14	63.2	0.204
	Yes	8	2	75.0	19	2	89.5	0.166	8	5	37.5	19	8	57.9	0.454
Body Mass Index	Healthy	9	2	77.8	31	2	93.5	0.260	9	4	55.6	31	7	77.4	0.354
	Obese	6	1	83.3	13	3	76.9	0.890	6	3	50.0	13	7	46.2	0.963
In situ component	No	18	2	88.9	53	8	84.9	0.745	18	8	55.6	53	18	66.0	0.483
	Yes	8	1	87.5	27	1	96.3	0.341	8	1	87.5	27	9	66.7	0.303
Histology	Ductal	22	3	86.4	56	6	89.3	0.766	22	8	63.6	56	19	66.1	0.749
	Not ductal	0	0	--	8	1	87.5	--	0	0	--	8	2	75.0	--
Histological Grade	Low/Inter ^c	19	2	89.5	53	3	94.3	0.494	19	6	68.4	53	17	67.9	0.868
	High	4	1	75.0	23	5	78.3	0.706	4	1	75.0	23	9	60.9	0.674

Table 14 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the variant allele of rs6493497

Covariables	Overall Survival						Progression Free Survival								
	Variant Allele			Wild-Type Allele			p ^a	Variant Allele			Wild-Type Allele			p ^a	
	n	Events	OS (%)	n	Events	OS (%)		n	Events	PFS (%)	n	Events	PFS (%)		
PgR Status															
Negative	5	1	80.0	14	3	78.6	0.963	5	2	60.0	14	4	71.4	0.527	
Positive	20	2	90.0	64	5	92.2	0.660	20	6	70.0	64	22	65.6	0.742	
HER2 Status															
Negative	4	2	50.0	12	3	75.0	0.214	4	3	25.0	12	7	41.7	0.457	
Positive	6	1	83.3	12	3	75.0	0.936	6	4	33.3	12	9	25.0	0.704	
Tumor size															
Small	12	0	100.0	33	3	90.9	0.294	12	4	66.7	33	8	75.8	0.801	
Large	13	2	84.6	47	6	87.2	0.735	13	4	69.2	47	19	59.6	0.535	
Lymph node involvement															
No	20	1	95.0	58	6	89.7	0.491	20	6	70.0	58	17	70.7	0.882	
Yes	5	1	80.0	21	3	85.7	0.363	5	2	60.0	21	10	52.4	0.915	
Distant Metastases															
Absent	25	2	92.0	73	8	89.0	0.739	25	8	68.0	73	22	69.9	0.962	
Present	1	1	0.0	5	1	80.0	0.025	1	1	0.0	5	4	20.0	0.583	
Tumor Stage															
Early	18	0	100.0	53	6	88.7	0.149	18	5	72.2	53	16	69.8	0.662	
Advanced	8	3	62.5	24	3	87.5	0.037	8	4	50.0	24	10	58.3	0.642	
Neoadjuvant Treatment															
No	26	3	88.5	70	8	88.6	0.905	26	9	65.4	70	23	67.1	0.970	
Yes	0	0	--	89	1	88.9	--	0	0	--	9	3	66.7	--	
Adjuvant Chemotherapy															
No	9	1	88.9	22	2	90.9	0.801	9	2	77.8	22	6	72.7	0.782	
Yes	17	2	88.2	58	7	87.9	0.924	17	7	58.8	58	21	63.8	0.817	
Adjuvant Radiotherapy															
No	6	2	66.7	11	2	81.8	0.397	6	3	50.0	11	5	54.5	0.719	
Yes	20	1	95.0	69	7	89.9	0.500	20	6	70.0	69	22	68.1	0.752	

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if p<0.050, the results are significantly different; ^b M – menopausal; ^c Inter. – Intermediate.

Concerning overall survival, we observed that in patients with distant metastases at diagnosis, carriers of the variant allele of rs6493497 had a significantly worse overall survival (0.0% over 80.0% for non-carriers, $p=0.025$). Also, carriers of the allele with advanced stage tumors had a worse overall survival than non-carriers (62.5% over 87.5%, $p=0.037$).

On the other hand, concerning progression free survival, we did not observe any significant difference between carriers and non-carriers of the allele.

In conclusion, we observed a tendency for the response to treatment to be worse for carriers of the A allele of rs6493497. A study [94] reports that the presence of the variant allele of this polymorphism is associated with higher basal estrogen levels, and this may explain the worse response to the treatment observed in carriers of this allele, because if estrogen levels are higher, there is more aromatase in the breast tissue, and more target to the aromatase inhibitors, rendering them less effective.

3.4. Tetranucleotide repetition polymorphism

We also performed survival analysis for the four most common alleles found in our population of the tetranucleotide repetition polymorphism: (TTTA)_{7(-3bp)}, (TTTA)₁₁, (TTTA)₇ and (TTTA)₈. The results will be discussed next.

3.4.1. (TTTA)_{7(-3bp)}

The results for overall and progression free survival are shown in figure 15. We did not observe any significant difference concerning overall (87.9% for non-carriers and 88.1% for carriers, $p=0.764$) and progression free survival (60.5% without and 69.5% with the allele, $p=0.345$).

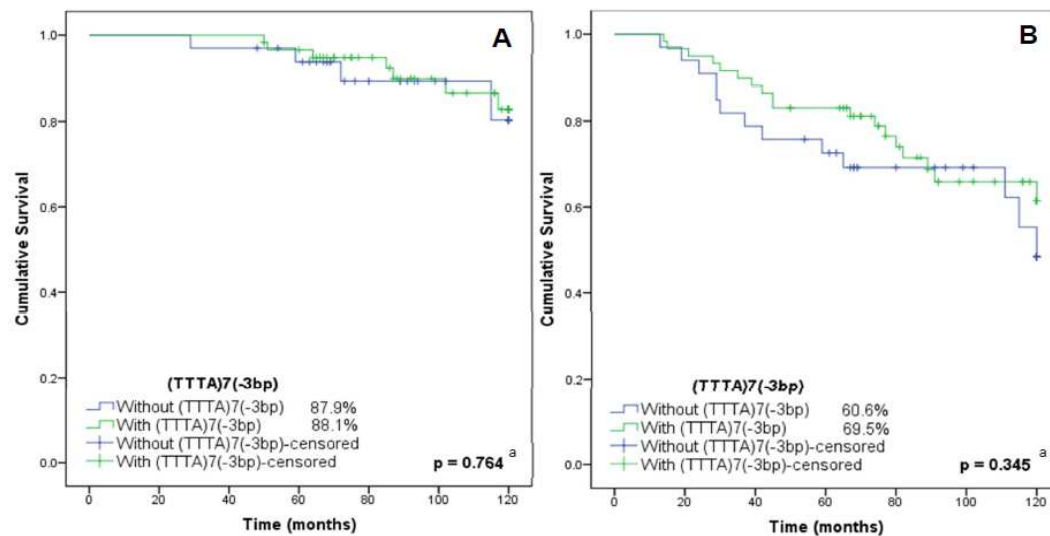


Figure 15: Overall and progression free survival for the presence or absence of the variant allele of (TTTA)_{7(-3bp)} obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

After stratification of the survival analyses for the parameters studied, we obtained the results shown in table 15.

Concerning overall survival, we did not observe any significant difference between carriers and non-carriers of the allele, in any of the subsets.

However, in relation to progression free survival, we observed that carriers of the allele with a family history of breast cancer had a significantly better progression free survival (73.3% compared to 14.3%, $p = 0.003$). We also observed a statistical trend for carriers of the allele who were diagnosed before 50 years of age to have a better progression free survival than non-carriers (63.6% for carriers over 36.4% for non-carriers, $p = 0.096$).

Table 15: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the (TTTA)₇(-3bp)

Covariables	Overall Survival						Progression Free Survival								
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele					
	n	Events	OS (%)	n	Events	OS (%)	p ^a	n	Events	PFS (%)	n	Events	PFS (%)	p ^a	
Age at Diagnosis (y)	>50	37	4	89.2	22	1	95.5	0.464	37	10	73.0	22	6	72.7	0.979
	<50	22	3	86.4	11	3	72.7	0.247	22	8	63.6	11	7	36.4	0.096
Menopausal Status	Post-M ^b	29	3	89.7	22	2	90.9	0.913	29	7	75.9	22	6	72.7	0.764
	Pre-M ^b	20	4	80.0	9	2	77.8	0.569	20	8	60.0	9	6	33.3	0.125
Age at Menarche	Late	25	4	84.0	15	2	86.7	0.717	25	9	64.0	15	5	66.7	0.792
	Early	21	3	85.7	8	1	87.5	0.955	21	6	71.4	8	4	50.0	0.487
Age at Menopause	Early	16	2	87.5	10	2	80.0	0.794	16	3	81.3	10	2	80.0	0.964
	Late	11	1	90.9	7	1	85.7	0.935	11	2	81.8	7	4	42.9	0.175
Breast Feeding	No	6	1	83.3	2	0	100.0	0.564	6	3	50.0	2	1	50.0	0.511
	Yes	9	1	88.9	10	2	80.0	0.692	9	2	77.8	10	4	60.0	0.625
Contraceptives	No	8	3	62.5	9	0	100.0	0.144	8	4	50.0	9	3	66.7	0.710
	Yes	18	2	88.9	9	3	66.7	0.170	18	7	61.1	9	6	33.3	0.145
Family history of breast cancer	No	27	4	85.2	17	3	82.4	0.831	27	10	63.0	17	6	64.7	0.916
	Yes	15	2	86.7	7	1	85.7	0.697	15	4	73.3	7	6	14.3	0.003
Body Mass Index	Healthy	22	2	90.9	13	2	84.6	0.470	22	6	72.7	13	5	61.5	0.475
	Obese	10	2	80.0	7	2	71.4	0.793	10	3	70.0	7	5	28.6	0.192
In situ component	No	39	6	84.6	23	3	87.0	0.875	39	12	69.2	23	12	47.8	0.152
	Yes	20	1	95.0	10	1	90.0	0.401	20	6	70.0	10	1	90.0	0.262
Histology	Ductal	44	5	88.6	23	4	82.6	0.462	44	14	68.2	23	10	56.5	0.307
	Not ductal	4	0	100.0	2	0	100.0	--	4	0	100.0	2	0	100.0	--
Histological Grade	Low/Inter ^c	40	2	95.0	23	2	91.3	0.430	40	10	75.0	23	9	60.9	0.225
	High	17	5	70.6	7	2	71.4	0.420	17	7	58.8	7	4	42.9	0.280

Table 15 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of the (TTTA)_{7(-3bp)}

Covariables	Overall Survival					Progression Free Survival				
	Variant Allele			Wild-Type Allele		P ^a	Variant Allele			P ^a
	n	Events	OS (%)	n	Events	OS (%)	n	Events	PFS (%)	PFS (%)
PgR Status										
Negative	13	3	76.9	3	0	100.0	13	4	69.2	100.0
Positive	46	4	91.3	29	4	86.2	46	14	69.6	55.2
										0.243
										0.207
HER2 Status										
Negative	10	2	80.0	5	2	60.0	5	5	0.0	60.0
Positive	12	3	75.0	5	2	60.0	12	8	33.3	20.0
										0.312
										0.332
Tumor size										
Small	24	2	91.7	16	0	100.0	24	5	79.2	62.5
Large	35	5	85.7	16	3	81.3	35	13	62.9	62.5
										0.915
Lymph node involvement										
No	43	5	88.4	28	2	92.9	43	12	72.1	64.3
Yes	15	2	86.7	4	1	75.0	15	6	60.0	50.0
										0.356
										0.872
Distant Metastases										
Absent	55	6	89.1	30	2	93.3	55	15	72.7	63.3
Present	2	1	50.0	3	2	33.3	2	2	0.0	33.3
										0.207
Tumor Stage										
Early	41	5	87.8	23	0	100.0	41	12	70.7	65.2
Advanced	15	2	86.7	10	4	60.0	15	5	66.7	50.0
										0.682
Neoadjuvant Treatment										
No	54	6	88.9	31	3	90.3	54	16	70.4	61.3
Yes	5	1	80.0	2	1	50.0	5	2	60.0	50.0
										0.329
										0.981
Adjuvant Chemotherapy										
No	13	1	92.3	13	1	92.3	13	2	84.6	76.9
Yes	46	6	87.0	20	3	85.0	46	16	65.2	50.0
										0.382
Adjuvant Radiotherapy										
No	5	1	80.0	8	1	87.5	5	2	60.0	75.0
Yes	54	6	88.9	25	3	88.0	54	16	70.4	56.0
										0.978
										0.267

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if p<0.050, the results are significantly different; ^bM – menopausal; ^cInter. – Intermediate.

So, according to our results, carriers of this allele show a tendency to have a better response to the treatment compared to non-carriers. A study [45] reports that the TCT deletion is associated with lower estrogen levels. This may explain the better response to the treatment observed in carriers of this allele in our study, as the one presented for the rs10046 polymorphism, because if there is less aromatase activity, there is less target for the inhibitor and higher effectiveness of the treatment.

3.4.2. (TTTA)₁₁

We analyzed the overall and progression free survival for the presence or absence of the (TTTA)₁₁ allele. The results are shown in figure 16. Here we observed a statistically significant difference in terms of overall survival, with carriers of the allele displaying a worse overall survival than non-carriers (77.4% compared to 93.4%, $p=0.011$). However, concerning progression free survival, we did not observe any significant difference (61.3% for carriers and 68.9% for non-carriers, $p=0.256$).

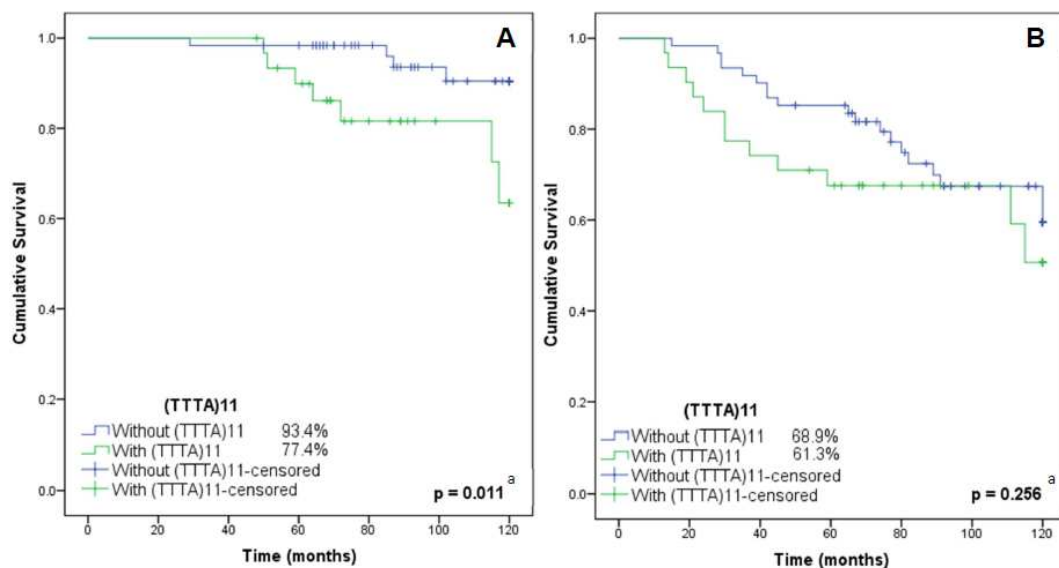


Figure 16: Overall and progression free survival for the presence or absence of the variant allele of (TTTA)₁₁ obtained in our population using the Kaplan-Meier survival analysis method. A correspond to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

Then we stratified the analysis once again to check if there is any effect for the allele in the different subsets of patients. The results are shown in table 16.

Table 16: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of (TTA)₁₁

Covariables	Overall Survival							Progression Free Survival						
	Variant Allele			Wild-Type Allele			P ^a	Variant Allele			Wild-Type Allele			P ^a
	n	Events	OS (%)	n	Events	OS (%)		n	Events	PFS (%)	n	Events	PFS (%)	
Age at Diagnosis (y)														
>50	18	2	88.9	41	3	92.7	0.470	18	3	83.3	41	13	68.3	0.345
<50	13	5	61.5	20	1	95.0	0.015	13	9	30.8	20	16	70.0	0.008
Menopausal Status														
Post-M ^b	18	2	88.9	33	3	90.9	0.766	18	3	83.3	33	10	69.7	0.316
Pre-M ^b	13	5	61.5	16	1	93.8	0.019	13	9	30.8	16	5	68.8	0.017
Age at Menarche														
Late	13	3	76.9	27	3	88.9	0.181	13	3	76.9	27	11	59.3	0.562
Early	11	3	72.7	18	1	94.4	0.063	11	6	45.5	18	4	77.8	0.127
Age at Menopause														
Early	5	1	80.0	21	3	85.7	0.908	5	1	80.0	21	4	81.0	0.961
Late	8	2	75.0	10	0	100.0	0.194	8	2	75.0	10	4	60.0	0.393
Breast Feeding														
No	1	0	100.0	7	1	85.9	--	1	0	100.0	7	4	42.9	0.475
Yes	8	2	75.0	11	1	90.9	0.405	8	3	62.5	11	3	72.7	0.663
Contraceptives														
No	7	1	85.7	10	2	80.0	0.968	7	1	85.7	10	6	40.0	0.169
Yes	9	4	55.6	18	1	94.4	0.018	9	7	22.2	18	6	66.7	0.004
Family history of breast cancer														
No	17	6	64.7	27	1	96.3	0.004	17	8	52.9	27	8	70.4	0.139
Yes	6	1	83.3	16	2	87.5	0.914	6	4	33.3	16	6	62.5	0.150
Body Mass Index														
Healthy	12	4	66.7	23	0	100.0	0.002	12	7	41.7	23	4	82.6	0.004
Obese	6	1	83.3	11	3	72.7	0.487	6	3	50.0	11	5	54.5	0.858
In situ component														
No	19	5	73.7	43	4	90.7	0.103	19	10	47.4	43	14	67.4	0.145
Yes	12	2	83.3	18	0	100.0	0.035	12	2	83.3	18	5	72.2	0.579
Histology														
Ductal	19	5	73.7	48	4	91.7	0.019	19	8	57.9	48	16	66.7	0.309
Not ductal	3	0	100.0	3	0	100.0	--	3	0	100.0	3	0	100.0	--
Histological Grade														
Low/Inter ^c	21	4	81.0	42	0	100.0	0.001	21	8	61.9	42	11	73.8	0.158
High	8	3	62.5	16	4	75.0	0.265	8	4	50.0	16	7	56.3	0.589

Table 16: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of (TTTA)¹¹

Covariables	Overall Survival						Progression Free Survival					
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele		
	n	Events	OS (%)	n	Events	OS (%)	n	Events	PFS (%)	n	Events	PFS (%)
PgR Status						<i>p^a</i>						<i>p^a</i>
Negative	3	1	66.7	13	2	84.6	3	1	66.7	13	3	76.9
Positive	27	6	77.8	48	2	95.8	27	11	59.3	48	16	66.7
						0.008						0.281
HER2 Status												
Negative	5	3	40.0	10	1	90.0	5	5	0.0	10	4	60.0
Positive	7	4	42.9	10	1	90.0	7	6	14.3	10	6	40.0
						0.245						0.053
						0.040						0.071
Tumor size												
Small	17	1	94.1	23	1	95.7	17	5	70.6	23	6	73.9
Large	14	6	57.1	37	2	94.6	14	7	50.0	37	12	67.6
						<0.001						0.144
Lymph node involvement												
No	26	5	80.8	45	2	95.6	26	10	61.5	45	12	73.3
Yes	4	2	50.0	15	1	93.3	4	2	50.0	15	6	60.0
						0.017						0.146
						0.096						0.684
Distant Metastases												
Absent	29	6	79.3	56	2	94.6	29	11	62.1	56	15	73.2
Present	2	1	50.0	3	2	33.3	2	1	50.0	3	3	0.0
						0.642						0.063
Tumor Stage												
Early	22	3	86.4	42	2	95.2	22	8	63.6	42	12	71.4
Advanced	8	4	50.0	17	2	88.2	8	4	50.0	17	6	64.7
						0.090						0.608
Neoadjuvant Treatment												
No	26	5	80.8	59	4	93.2	26	10	61.5	59	18	69.5
Yes	5	2	60.0	2	0	100.0	5	2	60.0	2	1	50.0
						0.032						0.195
						0.343						0.624
Adjuvant Chemotherapy												
No	10	1	90.0	16	1	93.8	10	1	90.0	16	4	75.0
Yes	21	6	71.4	45	3	93.3	21	11	47.6	45	15	66.7
						0.007						0.062
Adjuvant Radiotherapy												
No	5	0	100.0	8	2	75.0	5	1	80.0	8	3	62.5
Yes	26	7	73.1	53	2	96.2	26	11	57.7	53	16	69.8
						0.001						0.157

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if $p < 0.050$, the results are significantly different; ^b M – menopausal; ^c Inter. – Intermediate.

In the case of this allele, almost all of the subsets' survival was affected by its presence. Carriers of the allele who were diagnosed before 50 had a significantly lower overall survival than non-carriers (61.5%, compared to 95.0%, $p=0.015$), and lower progression free survival as well (30.8% over 70.0%, $p=0.008$). The same happens for carriers of the allele who were diagnosed before menopause onset, having a worse overall (61.5% for carriers over 93.8% for non-carriers, $p=0.019$) and progression free (30.8% for carriers and 68.8% for non-carriers, $p=0.017$) survival. Age at menarche and menopause and breast feeding were not significantly affected by the presence of the allele. On the other hand, women who used oral contraceptives and carried the allele had also a significantly worse overall (55.6% over 94.4%, $p=0.018$) and progression free (22.2% over 66.7%, $p=0.004$) survival than non-carriers. Family history of breast cancer was affected only in terms of overall survival, with carriers of the allele with no family history having a significantly worse overall survival than non-carriers (64.7% compared to 96.3%, $p=0.004$). Additionally, women with a healthy body mass index had also a significantly worse overall (66.7% over 100.0%, $p=0.002$) and progression free survival (41.7% over 82.6%, $p=0.004$). Also, carriers of the allele had a worse overall survival when *in situ* component was present (83.3% over 100.0%, $p=0.035$), the tumor was ductal in histology (73.7% over 91.7%, $p=0.019$), the histological grade was low to intermediate (81.0% over 100.0%, $p=0.001$) and the tumor was PgR-positive (77.8% over 95.8%) and HER2 positive (42.9% over 90.0%, $p=0.040$), with the progression free not being significantly affected in these subsets. In relation to the HER2 status, in terms of progression free survival we observed a non-significant trend for both patients with HER2-positive and -negative tumors to have a lower progression free survival (0.0% and 60.0%, $p=0.053$ for HER2-negative subset and 14.3% over 40.0%, $p=0.071$ for the HER2-positive subset). Overall survival was also worse for carriers of the allele with large tumors (57.1% against 94.6%, $p<0.001$) and low lymph node involvement (80.8% over 95.6%, $p=0.017$), with progression free survival not being affected. Concerning the presence of distant metastases, the results are interesting, because carriers of the allele without distant metastases had a worse overall survival (79.3% over 96.4%, $p=0.003$), whereas carriers with distant metastases had a better overall survival (50.0% over 33.3%, $p=0.642$). The same was observed non-significantly for progression free survival, in which carriers of the allele without distant metastases had a worse progression free survival (62.1% over 73.2%, $p=0.130$) and with distant metastases had a better progression free survival with a non-significant trend (50.0% over 0.0%, $p=0.063$). Concerning tumor stage we did not observe any significant

difference in terms of survival, except for a trend for carriers of the allele with both early and advanced stage breast cancer to have a lower overall survival (86.4% for early stage carriers over 95.2% for non-carriers, $p=0.097$, and 50.0% for advanced stage carriers over 88.2% for non-carriers, $p=0.090$). Progression free survival was not significantly affected in these subsets.

This allele also affected overall survival in patients who were not treated with neoadjuvant treatment (80.8% for carriers and 93.2% for non-carriers, $p=0.032$), but did not affect the progression free survival. Concerning patients treated with adjuvant chemotherapy, carriers had worse overall survival (71.4% over 93.3%, $p=0.007$) and a trend to worse progression free survival as well (47.6% over 66.7%, $p=0.062$). Finally, carriers who were treated with adjuvant radiotherapy had also a worse overall survival (73.1% over 96.2%) but no effect on progression free survival.

3.4.3. $(TTTA)_7$

Concerning the presence and absence of the $(TTTA)_7$ allele, the results are shown in figure 17.

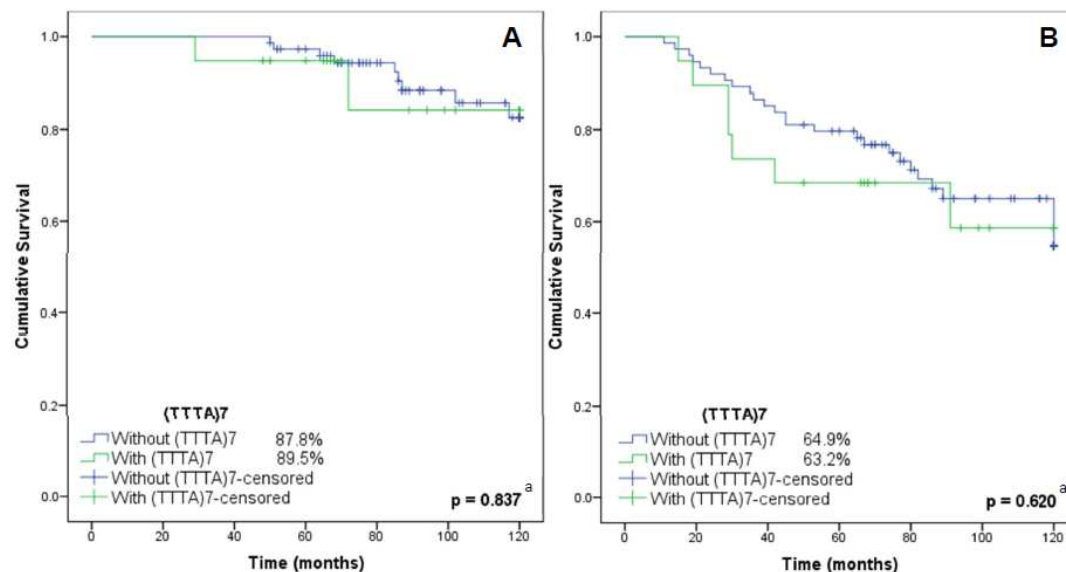


Figure 17: Overall and progression free survival for the presence or absence of the variant allele of $(TTTA)_7$ obtained in our population using the Kaplan-Meier survival analysis method. A corresponds to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

We did not observe any significant difference in terms of overall (89.5% for carriers and 87.9% for non-carriers, $p=0.837$) and progression free (63.2% for carriers and 64.9% for non-carriers, $p=0.620$) survival between carriers and non-carriers of this allele. This is not consistent with another study that reports a significant difference in both overall and progression free survival for carriers of this allele, who had a better outcome if the allele was present [100]. However, this difference can be due to the fact that in that study none of the patients was treated with aromatase inhibitors.

We then stratified the survival analysis to check if the survival was affected in any of the subsets between carriers and non-carriers of the allele. The results are shown in table 17.

In terms of overall survival, none of the subsets was significantly affected by the presence or absence of the (TTTA)₇ allele, except for a trend to carriers who breast fed to have a worse overall survival (66.7% for carriers and 93.8% for non-carriers, $p=0.090$), as well as carriers of the allele who ever used oral contraceptives to have a worse overall survival also (66.7% over 90.0%, $p=0.070$). A study reports a significantly better overall and progression free 8-year survival for premenopausal patients who carried the allele [100]. However, in this study no patients treated with aromatase inhibitors were included, and this might be the source of this difference.

Concerning progression free survival, carriers of the allele who had late menopause tend to have a lower progression free survival (25.0% over 69.2%, $p=0.074$). Also we observed that carriers of the allele with a family history of breast cancer had also a worse progression free survival (20.0% compared to 65.0%, $p=0.028$).

Table 17: Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of (TTTA)₇

Covariables	Overall Survival						Progression Free Survival								
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele					
	n	Events	OS (%)	n	Events	OS (%)	p ^a	n	Events	PFS (%)	n	Events	PFS (%)	p ^a	
Age at Diagnosis (y)	>50	13	0	100.0	47	5	89.4	0.304	13	4	69.2	47	15	68.1	0.751
	<50	6	2	66.7	27	4	85.2	0.166	6	3	50.0	27	11	59.3	0.471
Menopausal Status	Post-M ^b	13	1	92.3	28	4	89.5	0.948	13	4	69.2	28	11	71.1	0.718
	Pre-M ^b	4	1	75.0	33	4	82.6	0.468	4	2	50.0	33	9	60.9	0.458
Age at Menarche	Late	10	2	80.0	35	5	85.7	0.262	10	4	60.0	35	16	54.3	0.804
	Early	4	0	100.0	24	3	87.5	0.609	4	2	50.0	24	6	75.0	0.174
Age at Menopause	Early	9	2	77.8	19	3	84.2	0.509	9	2	77.8	19	5	73.7	0.891
	Late	4	0	100.0	13	1	92.3	0.579	4	3	25.0	13	4	69.2	0.074
Breast Feeding	No	2	0	100.0	6	1	83.3	0.617	2	1	50.0	6	3	50.0	0.511
	Yes	3	1	66.7	16	1	93.8	0.090	3	1	66.7	16	5	68.8	0.840
Contraceptives	No	3	0	100.0	14	3	78.6	0.395	3	2	33.3	14	6	57.1	0.391
	Yes	6	2	66.7	20	2	90.0	0.070	6	3	50.0	20	8	60.0	0.402
Family history of breast cancer	No	9	1	88.9	34	4	88.2	0.881	9	2	77.8	34	13	61.8	0.580
	Yes	5	1	80.0	20	3	85.0	0.210	5	4	20.0	20	7	65.0	0.028
Body Mass Index	Healthy	8	1	87.5	29	3	89.7	0.923	8	2	75.0	29	8	72.4	0.759
	Obese	5	1	80.0	10	2	80.0	0.157	5	3	40.0	10	5	50.0	0.192
In situ component	No	15	1	93.3	47	8	83.0	0.585	15	6	60.0	47	17	63.8	0.419
	Yes	4	1	75.0	27	1	96.3	0.106	4	1	75.0	27	9	66.7	0.805
Histology	Ductal	14	2	85.7	57	6	89.5	0.639	14	6	57.1	57	20	64.9	0.435
	Not ductal	1	0	100.0	7	1	85.7	0.683	1	0	100.0	7	2	71.4	0.558
Histological Grade	Low/Inter ^c	10	1	90.0	51	3	94.1	0.407	10	4	60.0	51	16	68.6	0.357
	High	6	1	83.3	21	5	76.2	0.940	6	2	66.7	21	8	61.9	0.965

Table 17 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of (TTTA)₇

Covariables	Overall Survival						Progression Free Survival								
	Variant Allele			Wild-Type Allele			p ^a	Variant Allele			Wild-Type Allele			p ^a	
	n	Events	OS (%)	n	Events	OS (%)		n	Events	PFS (%)	n	Events	PFS (%)		
PgR Status															
Negative	3	0	100.0	15	4	73.3	0.372	3	0	100.0	15	6	60.0	0.251	
Positive	16	2	87.5	57	4	93.0	0.301	16	7	56.3	57	18	68.4	0.267	
HER2 Status															
Negative	3	1	66.7	10	3	70.0	0.243	3	2	33.3	10	5	50.0	0.260	
Positive	4	1	75.0	14	3	78.6	0.455	4	2	50.0	14	11	21.4	0.881	
Tumor size															
Small	9	0	100.0	29	3	89.7	0.437	9	3	66.7	29	7	75.9	0.223	
Large	9	1	88.9	45	6	86.7	0.981	9	3	66.7	45	19	57.8	0.699	
Lymph node involvement															
No	16	1	93.8	52	6	88.5	0.658	16	5	68.8	52	16	69.2	0.882	
Yes	2	0	100.0	21	3	85.7	--	2	1	50.0	21	10	52.4	0.385	
Distant Metastases															
Absent	17	1	94.1	69	8	88.4	0.685	17	5	70.6	69	22	68.1	0.983	
Present	2	1	50.0	3	1	66.7	0.221	2	2	0.0	3	3	0.0	0.207	
Tumor Stage															
Early	13	0	100.0	50	6	88.0	0.255	13	4	69.2	50	15	70.0	0.766	
Advanced	6	2	66.7	21	3	85.7	0.163	6	3	50.0	21	10	52.4	0.713	
Neoadjuvant Treatment															
No	17	2	88.2	67	8	88.1	0.667	17	6	64.7	67	23	65.7	0.659	
Yes	2	0	100.0	6	1	83.3	0.564	2	1	50.0	6	2	66.7	0.511	
Adjuvant Chemotherapy															
No	6	1	83.3	20	2	90.0	0.544	6	2	66.7	20	6	70.0	0.615	
Yes	13	1	92.3	54	7	87.0	0.830	13	5	61.5	54	20	63.0	0.828	
Adjuvant Radiotherapy															
No	13	1	66.7	11	3	72.7	0.669	13	1	66.7	11	6	45.5	0.713	
Yes	6	1	93.8	63	6	90.5	0.906	6	6	62.5	63	20	68.3	0.484	

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if p<0.050, the results are significantly different; ^b M – menopausal; ^c Inter. – Intermediate.

3.4.4. $(TTTA)_8$

Finally, we studied the effect of the presence of the $(TTTA)_8$ allele in survival. The results are shown in figure 18.

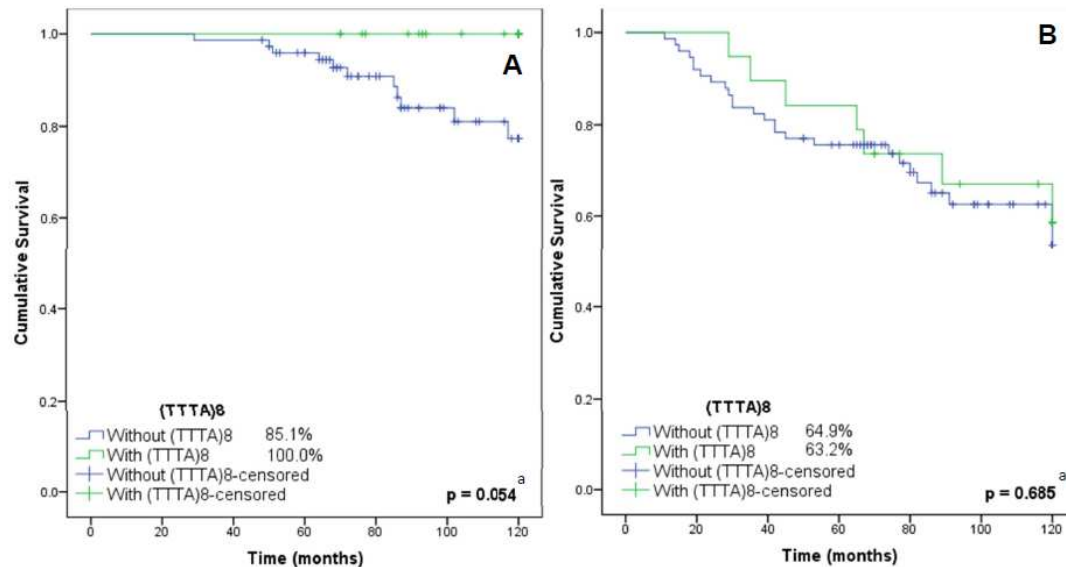


Figure 18: Overall and progression free survival for the presence or absence of the variant allele of $(TTTA)_8$ obtained in our population using the Kaplan-Meier survival analysis method. A corresponds to overall survival and B to progression free survival. ^alog-rank test in which if $p < 0.050$, the variables are significantly different.

Interestingly, none of the carriers of this allele died at the end of 120 months of follow-up (100.0% overall survival for carriers as opposed to 85.1% overall survival for non-carriers, $p=0.054$). On the other hand, in terms of progression free survival, we did not observe any significant difference between carriers and non-carriers of this allele (63.2% and 64.9%, respectively, $p=0.685$).

Considering the stratified analysis (table 18), carriers of the allele who did not receive any neoadjuvant treatment had a significantly better overall survival than non-carriers (100.0% over 84.6%, $p=0.045$). Besides, carriers of the allele with late menarche had a trend to better overall survival than non-carriers in the same subset (100.0% as opposed to 79.4%, $p=0.062$). Additionally, we also observed a trend for carriers of the allele with large tumors to have a better overall survival than non-carriers (100.0% over 82.9%, $p=0.091$). Finally, we also observed a statistical trend for carriers of the allele without distant metastases at diagnosis to have a better survival (100.0% over 87.0%, $p=0.097$).

Table 18. Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of (TTTA)₈

Covariables	Overall Survival						Progression Free Survival							
	Variant Allele			Wild-Type Allele			p ^a	Variant Allele			Wild-Type Allele			p ^a
	n	Events	OS (%)	n	Events	OS (%)		n	Events	PFS (%)	n	Events	PFS (%)	
Age at Diagnosis (y)														
>50	12	0	100.0	48	5	89.6	0.187	12	5	58.3	48	14	70.8	0.895
<50	7	0	100.0	26	6	76.9	0.163	7	2	71.4	26	12	53.8	0.391
Menopausal Status														
Post-M ^b	10	0	100.0	41	5	87.8	0.210	10	4	60.0	41	11	73.2	0.611
Pre-M ^b	5	0	100.0	22	5	77.3	0.281	5	2	60.0	22	9	59.1	0.939
Age at Menarche														
Late	11	0	100.0	34	7	79.4	0.062	11	6	45.5	34	14	58.8	0.992
Early	2	0	100.0	26	3	88.5	0.683	2	0	100.0	26	8	69.2	0.472
Age at Menopause														
Early	5	0	100.0	23	5	78.3	0.168	5	1	80.0	23	6	73.9	0.587
Late	4	0	100.0	13	1	92.3	0.579	4	2	50.0	13	5	61.5	0.856
Breast Feeding														
No	2	0	100.0	6	1	83.3	0.414	2	2	0.0	6	2	66.7	0.004
Yes	3	0	100.0	16	2	87.5	0.485	3	1	66.7	16	5	68.8	0.749
Contraceptives														
No	3	0	100.0	14	3	78.6	0.395	3	3	0.0	14	5	64.3	0.043
Yes	6	0	100.0	20	4	80.0	0.199	6	2	66.7	20	9	55.0	0.492
Family history of breast cancer														
No	8	0	100.0	35	5	85.7	0.221	8	3	62.5	35	12	65.7	0.895
Yes	6	0	100.0	19	4	78.9	0.176	6	3	50.0	19	8	57.9	0.934
Body Mass Index														
Healthy	5	0	100.0	32	4	87.5	0.345	5	1	80.0	32	9	71.9	0.570
Obese	1	0	100.0	14	3	78.6	0.789	1	0	100.0	14	8	42.9	0.513
In situ component														
No	11	0	100.0	51	9	82.4	0.118	11	4	63.6	51	19	62.7	0.610
Yes	8	0	100.0	23	2	91.3	0.346	8	3	62.5	23	7	69.6	0.934
Histology														
Ductal	14	0	100.0	57	8	86.0	0.181	14	5	64.3	57	21	63.2	0.718
Not ductal	2	0	100.0	6	1	83.3	0.527	2	0	100.0	6	2	66.7	0.362
Histological Grade														
Low/Inter ^c	16	0	100.0	45	4	91.1	0.163	16	6	62.5	45	14	68.9	0.887
High	2	0	100.0	25	6	76.0	0.435	2	1	50.0	25	9	64.0	0.841

Table 18 (continued): Overall and progression free survival analysis stratified to clinical and tumor characteristics and treatments applied for the presence or absence of (TTTA)₈

Covariables	Overall Survival						Progression Free Survival							
	Variant Allele			Wild-Type Allele			Variant Allele			Wild-Type Allele				
	n	Events	OS (%)	n	Events	OS (%)	p ^a	n	Events	PFS (%)	n	Events	PFS (%)	p ^a
PgR Status														
Negative	1	0	100.0	17	4	76.5	0.553	1	0	100.0	17	6	64.7	0.484
Positive	18	0	100.0	55	6	89.1	0.112	18	7	61.1	55	18	67.3	0.930
HER2 Status														
Negative	1	0	100.0	12	4	66.7	0.365	1	0	100.0	12	7	41.7	0.214
Positive	5	0	100.0	13	4	69.2	0.138	5	4	20.0	13	9	30.8	0.976
Tumor size														
Small	6	0	100.0	32	3	90.6	0.397	6	3	50.0	32	7	78.1	0.168
Large	13	0	100.0	41	7	82.9	0.091	13	4	69.2	41	18	56.1	0.224
Lymph node involvement														
No	14	0	100.0	54	7	87.0	0.127	14	5	64.3	54	16	70.4	0.859
Yes	5	0	100.0	18	3	83.3	0.284	5	2	60.0	18	9	50.0	0.351
Distant Metastases														
Absent	17	0	100.0	69	9	87.0	0.097	17	6	64.7	69	21	69.6	0.985
Present	0	0	100.0	5	2	60.0	--	0	0	--	5	5	0.0	--
Tumor Stage														
Early	12	0	100.0	51	6	88.2	0.194	12	5	58.3	51	14	72.5	0.442
Advanced	5	0	100.0	22	5	77.3	0.217	5	1	80.0	22	12	45.5	0.135
Neoadjuvant Treatment														
No	19	0	100.0	65	10	84.6	0.045	19	7	63.2	65	22	66.2	0.679
Yes	0	0	100.0	8	1	87.5	--	0	0	--	8	3	62.5	--
Adjuvant Chemotherapy														
No	6	0	100.0	20	3	85.0	0.268	6	2	66.7	20	6	70.0	0.902
Yes	13	0	100.0	54	8	85.2	0.123	13	5	61.5	54	20	63.0	0.761
Adjuvant Radiotherapy														
No	3	0	100.0	11	4	63.6	0.174	3	0	100.0	11	7	36.4	0.071
Yes	16	0	100.0	63	7	88.9	0.132	16	7	56.3	63	19	69.8	0.660

^a Overall and progression free survival obtained by Kaplan-Meier analysis, using a log-rank test in which if p<0.050, the results are significantly different, ^b M – menopausal, ^c Inter. – Intermediate.

On the other hand, considering progression free survival, carriers of the allele who never breast fed had a significantly worse progression free survival (0.0% over 66.7%, $p=0.004$) and the same was observed for carriers who never used oral contraceptives (0.0% over 64.3%, $p=0.043$). Finally, we observed a non-significant trend for carriers of the allele who did not receive adjuvant radiotherapy to have a better progression free survival than non-carriers (100.0% compared to 36.4%, $p=0.071$).

4. *Linkage Disequilibrium Analysis*

Throughout the analysis of our results, we observed some similarities in the behavior of some of the polymorphisms, especially between rs10046 and rs700518, which we found had the same minor allele frequency (48.2% for both, table 5), and both of them significantly improve overall survival in carriers with not ductal carcinomas and progression free survival in carriers with healthy body mass index (tables 12 and 13). Additionally, some studies report linkage disequilibrium relationships between rs10046 and (TTTA) $_n$ [92, 102]. For that reason we did a linkage disequilibrium analysis for all the alleles present in this study, to determine if the aforementioned associations are confirmed and if we find any new data on that issue.

The results for linkage disequilibrium analysis among the SNPs are shown in table 19. We observed that rs4646 is in strong linkage disequilibrium with rs10046 and rs700518. Also, we confirmed that in fact the rs10046 SNP is in strong linkage disequilibrium with rs700518, which explains the similarities between these two polymorphisms. However, rs6493497 is not in linkage disequilibrium with any of the other SNPs.

Table 19: Linkage disequilibrium analysis performed for the four SNPs included in our study

rs4646																
rs10046																
rs700518																
rs6493497																
	GG	GA	AA		AA	AG	GG		CC	CT	TT		GG	GA	AA	
	n (%)	n (%)	n (%)	<i>P</i> ^a	n (%)	n (%)	n (%)	<i>P</i> ^a	n (%)	n (%)	n (%)	<i>P</i> ^a	n (%)	n (%)	n (%)	<i>P</i> ^a
rs4646	GG				29	25	7		28	26	7		37	20	2	
					(26.6)	(22.9)	(6.4)		(25.6)	(23.9)	(6.4)		(34.9)	(18.9)	(1.9)	
					0	30	12	<0.001	1	29	12	<0.001	38	3	0	0.128
	GA				(0.0)	(27.5)	(11.0)		(0.9)	(26.6)	(11.0)		(35.8)	(2.8)	(0.0)	
	AA				0	0	6		0	0	6		5	1	0	
					(0.0)	(0.0)	(5.5)		(0.0)	(0.0)	(5.5)		(4.7)	(0.9)	(0.0)	
rs10046	AA				29	0	0		25	4	0		17	10	0	
					(26.6)	(0.0)	(0.0)		(22.9)	(3.7)	(0.0)		(16.0)	(9.4)	(0.0)	
					25	30	0	<0.001	4	49	2	<0.001	43	9	2	0.858
	AG				(22.9)	(27.5)	(0.0)		(3.7)	(45.0)	(1.8)		(40.6)	(8.5)	(1.9)	
	GG				7	12	6		0	2	23		20	5	0	
					(6.4)	(11.0)	(5.5)		(0.0)	(1.8)	(21.1)		(18.9)	(4.7)	(0.0)	
rs700518	CC				28	1	0		25	4	0		20	7	0	
					(25.6)	(0.9)	(0.0)		(22.9)	(3.7)	(0.0)		(18.9)	(6.6)	(0.0)	
					26	29	0	<0.001	4	49	2	<0.001	40	12	1	0.996
	CT				(23.9)	(26.6)	(0.0)		(3.7)	(45.0)	(1.8)		(37.7)	(11.3)	(0.9)	
	TT				7	12	6		0	2	23		20	5	0	
					(6.4)	(11.0)	(5.5)		(0.0)	(1.8)	(21.1)		(18.9)	(4.7)	(0.0)	
rs6493497	GG				37	38	5		17	43	20		20	40	20	
					(34.9)	(35.8)	(4.7)		(16.0)	(40.6)	(18.9)		(18.9)	(37.7)	(18.9)	
					20	3	1	0.128	10	9	5	0.858	7	12	5	0.995
	GA				(18.9)	(2.8)	(0.9)		(9.4)	(8.5)	(4.7)		(6.6)	(11.3)	(4.7)	
	AA				2	0	0		0	2	0		0	1	0	
					(1.9)	(0.0)	(0.0)		(0.0)	(1.9)	(0.0)		(0.0)	(0.9)	(0.0)	

^a Chi-Square test for association between the different SNPs, in which if $p < 0.050$, the variables are not independent.

Concerning the relationship between the SNPs and the tetranucleotide repetition polymorphism, the results are shown in table 20.

We observed that the rs4646 SNP is in linkage disequilibrium with the alleles of (TTTA)_n containing seven, with the TCT deletion and eleven repetitions, as well as a suggestive association with the seven repetition allele. On the other hand, rs10046 is in linkage only with the eleven repetition allele. A study reports a strong linkage disequilibrium relationship between the (TTTA)₇ allele and the G allele of rs10046 [92]. We did not observe that relationship. The most likely reason for that is the fact that in that study the authors did not discriminate between the seven repetition allele with and without the TCT deletion, so probably the linkage association they mention is biased in that way. Yet another study reports linkage disequilibrium between more than 8 repetitions in (TTTA)_n and the A allele of rs10046 [102]. In our study we only observed a relationship between this allele and the eleven repetition allele. The reason for that may be the low frequency of the alleles containing nine, ten and twelve repetitions, which if they were higher maybe we would observe a significant association. Then again, we observed linkage disequilibrium between the rs700518 and the seven repetition allele with the TCT deletion and the eleven repetition allele. Finally, we did not find any significant association between the rs6493497 SNP and any of the tetranucleotide repetition alleles.

It would be interesting, considering the obtained results, to do a haplotype analysis, in order to assess the combined effect of the studied polymorphisms, in terms of associations with clinical and tumor features and survival. However, unfortunately this was not possible because if we created an haplotype, our population size, which is already small, would be even more restricted and it would not be possible to find any significant associations.

Table 20: Linkage disequilibrium analysis performed for the four SNPs and the tetranucleotide repetition polymorphism

	rs4646				rs10046				rs700518				rs6493497			
	GG	GA	AA	P^a	AA	AG	GG	P^a	CC	CT	TT	P^a	GG	GA	AA	P^a
	n (%)	n (%)	n (%)		n (%)	n (%)	n (%)		n (%)	n (%)	n (%)		n (%)	n (%)	n (%)	
(TTTA) ₇																
AA ^b	0 (0.0)	2 (2.2)	1 (1.1)		1 (1.1)	2 (2.2)	0 (0.0)		1 (1.1)	2 (2.2)	0 (0.0)		0 (0.0)	0 (0.0)	3 (3.3)	
AB ^b	5 (5.4)	11 (11.8)	1 (1.1)	0.064	6 (6.4)	10 (10.7)	1 (1.1)	0.399	7 (7.5)	10 (10.7)	0 (0.0)	0.103	0 (0.0)	7 (7.8)	10 (11.1)	0.999
BB ^b	44 (43.7)	25 (26.9)	4 (4.3)		18 (19.4)	34 (36.6)	21 (22.6)		16 (16.7)	35 (37.6)	22 (23.7)		2 (2.2)	15 (16.7)	53 (58.9)	
(TTTA) _{7(-3bp)}																
AA ^b	4 (4.3)	5 (5.4)	6 (6.4)		14 (15.0)	1 (1.1)	0 (0.0)		14 (15.0)	1 (1.1)	0 (0.0)		0 (0.0)	4 (4.4)	11 (12.2)	
AB ^b	1 (1.1)	25 (26.9)	18 (19.4)	0.010	10 (10.7)	32 (34.4)	2 (2.2)	0.131	9 (9.7)	34 (36.6)	1 (1.1)	<0.001	2 (2.2)	7 (7.8)	34 (37.8)	0.989
BB ^b	1 (1.1)	8 (8.6)	25 (26.9)		1 (1.1)	13 (14.0)	20 (21.5)		1 (1.1)	12 (12.9)	21 (22.6)		0 (0.0)	11 (12.2)	21 (23.3)	
(TTTA) ₈																
AA ^b	0 (0.0)	0 (0.0)	1 (1.1)		0 (0.0)	0 (0.0)	1 (1.1)		0 (0.0)	0 (0.0)	1 (1.1)		0 (0.0)	0 (0.0)	1 (1.1)	
AB ^b	0 (0.0)	14 (15.0)	4 (4.3)	0.528	2 (2.2)	16 (17.2)	0 (0.0)	0.995	2 (2.2)	16 (17.2)	0 (0.0)	0.999	0 (0.0)	2 (2.2)	16 (17.8)	0.423
BB ^b	6 (6.4)	24 (25.8)	44 (47.3)		23 (24.7)	30 (32.3)	21 (22.6)		22 (23.7)	31 (33.3)	21 (22.6)		2 (2.2)	20 (22.2)	49 (54.4)	
(TTTA) ₉																
AA ^b	0 (0.0)	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)	
AB ^b	0 (0.0)	2 (2.2)	12 (12.9)	0.171	2 (2.2)	10 (10.7)	2 (2.2)	0.999	0 (0.0)	13 (14.0)	1 (1.1)	0.579	2 (2.2)	3 (3.3)	9 (10.0)	0.779
BB ^b	6 (6.4)	36 (41.9)	37 (39.8)		23 (24.7)	36 (41.9)	20 (21.5)		23 (24.7)	34 (36.6)	21 (22.6)		0 (0.0)	19 (21.1)	57 (63.3)	

Table 20 (continued): Linkage disequilibrium analysis performed for the four SNPs and the tetranucleotide repetition polymorphism

	rs4646					rs10046					rs700518					rs6493497				
	GG n (%)	GA n (%)	AA n (%)	<i>p</i> ^a	AA n (%)	AG n (%)	GG n (%)	CC n (%)	CT n (%)	TT n (%)	<i>p</i> ^a	GG n (%)	GA n (%)	AA n (%)	<i>p</i> ^a	GG n (%)	GA n (%)	AA n (%)	<i>p</i> ^a	
(TTTAA) ₁₀	AA ^b	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	0 (0.0)		
	AB ^b	0 (0.0)	1 (1.1)	4 (4.3)	0.773	0 (0.0)	3 (3.2)	2 (2.2)	0 (0.0)	3 (3.2)	0.517	0 (0.0)	3 (3.3)	2 (2.2)	0.536	0 (0.0)	3 (3.3)	2 (2.2)	0.689	
	BB ^b	6 (6.4)	37 (39.8)	45 (48.4)		25 (26.9)	43 (46.2)	20 (21.5)	24 (25.8)	44 (47.3)		2 (2.2)	19 (21.1)	64 (71.1)		2 (2.2)	19 (21.1)	64 (71.1)		
(TTTA) ₁₁	AA ^b	0 (0.0)	0 (0.0)	16 (17.2)		0 (0.0)	1 (1.1)	15 (16.1)	0 (0.0)	0 (0.0)		0 (0.0)	5 (5.6)	9 (10.0)		0 (0.0)	5 (5.6)	9 (10.0)		
	AB ^b	0 (0.0)	8 (8.6)	7 (7.5)	<0.001	0 (0.0)	10 (10.7)	5 (5.4)	0 (0.0)	11 (11.8)	<0.001	0 (0.0)	4 (4.4)	11 (12.2)	<0.001	0 (0.0)	4 (4.4)	11 (12.2)	0.996	
	BB ^b	6 (6.4)	30 (32.3)	26 (28.0)		25 (26.9)	35 (37.6)	2 (2.2)	24 (25.8)	36 (41.9)		2 (2.2)	13 (14.4)	46 (51.1)		2 (2.2)	13 (14.4)	46 (51.1)		
(TTTA) ₁₂	AA ^b	0 (0.0)	0 (0.0)	1 (1.1)		0 (0.0)	1 (1.1)	0 (0.0)	0 (0.0)	1 (1.1)		0 (0.0)	0 (0.0)	1 (1.1)		0 (0.0)	0 (0.0)	1 (1.1)		
	AB ^b	0 (0.0)	1 (1.1)	0 (0.0)	0.973	0 (0.0)	1 (1.1)	0 (0.0)	0 (0.0)	1 (1.1)	1.000	0 (0.0)	0 (0.0)	0 (0.0)	0.566	0 (0.0)	0 (0.0)	0 (0.0)	0.960	
	BB ^b	6 (6.4)	37 (39.8)	48 (51.6)		25 (26.9)	44 (47.3)	22 (23.7)	24 (25.8)	46 (49.5)		2 (2.2)	22 (24.4)	65 (72.2)		2 (2.2)	22 (24.4)	65 (72.2)		

^a Chi-Square test for association between the different SNPs, in which if $p < 0.050$, the variables are not independent. ^b A consists of the referred allele and B of any other allele.

5. Overall Discussion

In summary, considering all the results obtained in this study, we found some interesting associations between the different polymorphisms studied and clinical features and survival.

Concerning the rs4646 polymorphism, we observed a suggestive association between the A (variant) allele and healthy body mass index. Also, despite not affecting the overall survival, this allele is significantly associated with a worse progression free survival in carriers who never used oral contraceptives and had invasive carcinomas which were not ductal in histology, but had an *in situ* component associated.

On the other hand, we observed strong linkage disequilibrium between the rs10046 and rs700518 polymorphism, and very similar results between these two. Patients with carcinomas which were invasive but not ductal had a significantly better overall survival if the variant allele of rs10046 or rs700518 were present, although these alleles were not overrepresented in any clinical or tumor feature from the ones analyzed. There was however a difference in progression free survival between these two alleles, because while carriers of the G (variant) allele of rs10046 had no difference in progression free survival as compared to non-carriers, carriers of the T (variant) allele of rs700518 with healthy body mass index had a significantly better progression free survival than non-carriers.

For the rs6493497 polymorphism, we observed that the variant allele (A) was not significantly associated with diagnosis after menopause and the presence of invasive ductal carcinomas. Additionally, carriers of this rs6493497 allele who had distant metastases and advanced stage tumors had a significantly worse overall survival than non-carriers, while no effect was observed concerning progression free survival.

The tetranucleotide repetition polymorphism was also studied. We considered four of the alleles because they were the most frequent and more likely for an association to be observed. However, ideally we would prefer to analyze all of the possible alleles of this polymorphism. The four alleles studied were the seven repetition allele, with and without the TCT deletion, the eleven repetition allele and the eight repetition allele.

Concerning the seven repetition allele with TCT deletion, we did not observe any overrepresentation of this on any of the evaluated clinical and tumor features, and neither did we observe any association with overall survival. However, we

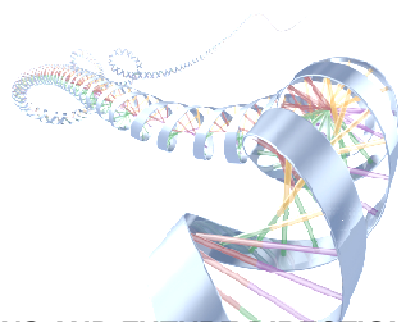
observed that carriers of this allele who were diagnosed before 50 years of age had a suggestive better progression free survival than non-carriers. Furthermore, we observed that carriers with a family history of breast cancer had a significantly better progression free survival than non-carriers.

On the other hand, concerning the eleven repetition allele, we observed that carriers were more likely to have a late menopause. Concerning overall survival, we can almost define a complete profile of overall survival for carriers of this allele: patients with early diagnosis and pre-menopausal at diagnosis, early menarche, no family history of breast cancer, healthy weight, and with large invasive ductal carcinomas with an in situ component associated, of low to intermediate histological grade, which are PgR and HER2 positive, with low number of metastases in the lymph nodes and no distant metastases, of any stage, and who were not administered neoadjuvant treatment, but were treated with adjuvant chemotherapy and radiotherapy have a worse overall survival if the eleven repetition allele is present. Also, carriers of this allele had a worse progression free survival if they were diagnosed before 50 years of age and had a healthy body mass index, and may have a worse progression free survival as well if distant metastases are present at diagnosis and were subjected to adjuvant chemotherapy. This allele is worth further study, with a larger population and better defined clinical, tumor and treatment features, because it is very likely to be a predictor of worse treatment response.

For the seven repetition allele, without TCT deletion, although it was not overrepresented in any of the clinical and tumor features evaluated, carriers may have a worse overall survival if they breast fed or used oral contraceptives. Also, carriers of this allele who had late menopause may have a worse progression free survival and a significantly worse progression free survival if a family history of breast cancer is observed.

Finally, carriers of the eight repetition allele are more likely to have a late onset of menarche and PgR positive tumors with low to intermediate histological grade. Besides, carriers of this allele who had large tumors with no distant metastases at diagnosis may have a better overall survival than non-carriers, and significantly better overall survival if they do not receive neo-adjuvant treatment. Concerning progression free survival, it was worse if carriers did not breast feed and never used oral contraceptives and may be better if carriers were treated with adjuvant radiotherapy.

Some studies report that estrogen levels may explain the differences in treatment responses, mentioning that the higher the levels of the hormone, the better is the response, because the decrease in estrogen levels is higher upon aromatase inhibitor administration [29, 91, 100]. However, although we were not able to measure estrogen levels in our study population, we verified that based on the results from other studies, the response is consistently worse when estrogen levels are reportedly increased. So we propose an alternative explanation for that: if the estrogen levels are high, there is more aromatase in the breast tissue, and more substrate for the aromatase inhibitors. Therefore, aromatase inhibitors are not able to interact as effectively with the enzyme. This proposition is supported by another study, that suggests the same explanation [94]. However, this is just speculation, because we cannot exclude the limitations inherent to this study, such as the small study population size, the lack of information concerning some parameters and the absence of hormone level measurements and functional analyses. Additionally, the exact mechanisms underlying the effects of *CYP19A1* polymorphisms on treatment response are not known, and these differences may also be due to linkage disequilibrium between the alleles included in our study with other functional alleles, as well as the fact that these polymorphisms may have an effect on other genes or other proteins and affect in this way the breast cancer survival and the response to the treatment.

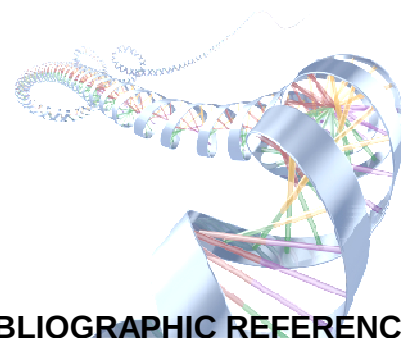


V. CONCLUSIONS AND FUTURE DIRECTIONS

In conclusion, our results show that there is an effect of *CYP19A1* polymorphisms on breast cancer treatment response. Concerning the SNPs, the presence of the variant allele of rs4646 is suggestive of a worse response, as well as the presence of the variant allele of rs6493497. On the other hand, the carriers of the variant allele of rs10046 or rs700518 probably have a better response to treatment with aromatase inhibitors. Additionally, we also observed that the tetranucleotide repetition polymorphism also affects the response to the treatment, especially the presence of the eleven repetition allele, which conferred a significantly worse response, in terms of overall as well as progression free survival. The eight repetition allele also originated interesting results, showing that carriers of this allele do not die after ten years of follow-up, resulting in a better response to the treatment. Finally, concerning the seven repetition allele, the presence of a 3bp deletion associated with this allele suggestively confers a better response than its absence.

However, all of these results need further confirmation, considering the limitations of this work, such as small sample size and us not being able to collect some data from all of the patients, as well as not being able to measure hormone levels. Therefore, in the future, it would be interesting to increase the sample size for us to be capable to observe stronger associations, not only in terms of survival, but also in terms of the effect that the presence of the studied alleles has on patient and tumor features. Also, a larger sample size would allow us to study all the alleles of the tetranucleotide repetition polymorphism, as well as to do a haplotype analysis, in order to study the combined effect of these polymorphisms on treatment response. If the population size was larger it would also be possible to perform a multivariate analysis, instead of a univariate analysis such as the one we did, allowing us to assess the effects of the polymorphisms on survival stratified for multiple variables at the same time. Additionally, it would also be necessary to do functional assays concerning these polymorphisms, in order to assess the true functional effects inherent to them, as well as measurements of hormone levels, preferably before and after the treatment, to assess whether the presence of any of these alleles has a bigger impact on hormone level decrease after treatment, and also to find out the true role of hormone levels in the response to the treatment. It would also be interesting to study a larger number of polymorphisms, so that we could find any functional polymorphism, or a polymorphism with even stronger effects than the ones observed.

Nevertheless, this is one of the few studies already published to assess the effects of the *CYP19A1* on breast cancer treatment response, and the only one to do a complete analysis of the association with inherent patient and tumor characteristics as well as a survival analysis considering so many potential prognostic factors. Our results further confirm the importance of this kind of studies and the need for more of these studies in order to increase the treatment effectiveness and individualization, to reduce the risk of absence of response and therefore improve the patients survival to breast cancer, as well as their quality of life.



VI. BIBLIOGRAPHIC REFERENCES

1. Mitrunen, K. and A. Hirvonen, *Molecular epidemiology of sporadic breast cancer: The role of polymorphic genes involved in oestrogen biosynthesis and metabolism*. Mutation Research, 2003. **544**: p. 33.
2. Kristensen, V.N. and A.L. Børresen-Dale, *Molecular epidemiology of breast cancer: genetic variation in steroid hormone metabolism*. Mutation Research, 2000. **462**: p. 11.
3. Bartsch, H., J. Nair, and R. Owen, *Dietary polyunsaturated fatty acids and cancers of the breast and colorectum: emerging evidence for their role as risk modifiers* Carcinogenesis, 1999. **20**: p. 10.
4. Zheng, W., et al., *Well-done meat intake and the risk of breast cancer* Journal of the National Cancer Institute, 1998. **90**: p. 6.
5. McKeown, N., *Antioxidants and breast cancer*. Nutrition Reviews, 1999. **57**: p. 4.
6. Hunter, D. and W. Willett, *Diet, body size, and breast cancer*. Epidemiology Reviews, 1993. **15**(1): p. 23.
7. Männistö, S., et al., *Body-size indicators and risk of breast cancer according to menopause and estrogen-receptor status*. International Journal of Cancer, 1996. **68**(1): p. 6.
8. Henderson, B., et al., *Do regular ovulatory cycles increase breast cancer risk?* Cancer, 1985. **56**(5): p. 3.
9. Olson, J.E., et al., *A Comprehensive Examination of CYP19 Variation and Risk of Breast Cancer Using Two Haplotype Tagging Approaches*. Breast Cancer Research and Treatment, 2007. **102**(2): p. 18.
10. Palmer, J. and L. Rosenberg, *Cigarette smoking and the risk of breast cancer*. Epidemiology Reviews, 1993. **15**(1): p. 12.
11. Baron, J., C.L. Vecchia, and F. Levi, *The antiestrogenic effect of cigarette smoking in women*. American Journal of Obstetrics and Gynecology, 1990. **162**(2): p. 13.
12. Kuper, H., et al., *Alcohol and breast cancer risk: the alcoholism paradox*. British Journal of Cancer, 2000. **83**(7): p. 3.
13. Reichman, M., et al., *Effects of alcohol consumption on plasma and urinary hormone concentrations in premenopausal women*. Journal of the National Cancer Institute, 1993. **85**(9): p. 6.

14. Wright, R., J. McManaman, and J. Repine, *Alcohol-induced breast cancer: a proposed mechanism*. Free Radical Biology and Medicine, 1999. **26**(3-4): p. 7.
15. Barnes, S., K. Singletary, and R. Frey, *Ethanol and acetaldehyde enhance benzo[a]pyrene-DNA adduct formation in human mammary epithelial cells*. Carcinogenesis, 2000. **21**(11): p. 6.
16. Cancer, C.G.o.H.F.i.B., *Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies*. Lancet, 1996. **347**(9017): p. 15.
17. Holli, K., J. Isola, and J. Cuzick, *Hormone replacement therapy and biological aggressiveness of breast cancer*. Lancet, 1997. **350**(9092): p. 2.
18. Haiman, C.A., et al., *A comprehensive haplotype analysis of CYP19 and breast cancer risk: the Multiethnic Cohort*. Human Molecular Genetics, 2003. **12**(20): p. 14.
19. Dorgan, J.F., et al., *Serum Sex Hormone Levels Are Related to Breast Cancer Risk in Postmenopausal Women*. Environmental Health Perspectives, 1997. **105**(3): p. 3.
20. Warren, R., et al., *Associations among Mammographic Density, Circulating Sex Hormones, and Polymorphisms in Sex Hormone Metabolism Genes in Postmenopausal Women*. Cancer Epidemiology Biomarkers and Prevention, 2006. **15**(8): p. 7.
21. Boyd, N., et al., *Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study*. Journal of the National Cancer Institute, 1995. **87**(9): p. 6.
22. Boyd, N., et al., *A longitudinal study of the effects of menopause on mammographic features*. Cancer Epidemiology Biomarkers and Prevention, 2002. **11**(10): p. 6.
23. Cuzick, J., et al., *Tamoxifen and breast density in women at increased risk of breast cancer*. Journal of the National Cancer Institute, 2004. **96**(8): p. 8.
24. Rutter, C., et al., *Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy*. Journal of the American Medical Association, 2001. **285**(2): p. 6.
25. Jonh, E. and J. Kelsey, *Radiation and other environmental exposures and breast cancer*. Epidemiology Reviews, 1993. **15**: p. 6.
26. Pukkala, E. and F. Weiderpass, *Time trends in socio-economic differences in incidence rates of cancers of the breast and female genital organs*. International Journal of Cancer, 1999. **81**: p. 6.

-
27. Garner, C., H. Matthews, and L. Burka, *Phenolphthalein metabolite inhibits catechol-O-methyltransferase-mediated metabolism of catechol estrogens: a possible mechanism for carcinogenicity*. Toxicology and Applied Pharmacology, 2000. **162**(2): p. 8.
 28. Ribeiro, F.S., et al., *CYP19 (TTTA)*n* polymorphism and breast cancer risk in Brazilian women*. Toxicology Letters, 2006. **164**: p. 6.
 29. Fasching, P.A., et al., *Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients*. Breast Cancer Research and Treatment, 2008. **112**: p. 10.
 30. Key, T., et al., *Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies*. Journal of the National Cancer Institute, 2002. **94**(8): p. 16.
 31. Miyoshi, Y., et al., *Association of genetic polymorphisms in CYP19 and CYP1A1 with the oestrogen receptor-positive breast cancer risk*. European Journal of Cancer, 2003. **39**: p. 7.
 32. Okobia, M.N., et al., *Simple tandem repeat (TTTA)*n* polymorphism in CYP19 (aromatase) gene and breast cancer risk in Nigerian women*. Journal of Carcinogenesis, 2006. **5**(12): p. 8.
 33. Conley, A. and M. Hinshelwood, *Mammalian aromatases*. Reproduction, 2001. **121**: p. 11.
 34. Bulun, S.E., et al., *The human CYP19 (aromatase P450) gene: update on physiologic roles and genomic organization of promoters*. Journal of Steroid Biochemistry & Molecular Biology, 2003. **86**: p. 6.
 35. Fabian, C., *The what, why and how of aromatase inhibitors: hormonal agents for treatment and prevention of breast cancer*. International Journal of Clinical Practice, 2007. **61**(12): p. 13.
 36. Liehr, J., *Is estradiol a genotoxic mutagenic carcinogen?* Endocrinology Reviews, 2000. **21**(1): p. 15.
 37. Simpson, E.R. and S.R. Davis, *Minireview: Aromatase and the Regulation of Estrogen Biosynthesis—Some New Perspectives*. Endocrinology, 2001. **142**(11): p. 7.
 38. Czajka-Oraniec, I. and E.R. Simpson, *Aromatase research and its clinical significance*. Polish Journal of Endocrinology, 2010. **61**(1): p. 9.
 39. Simpson, E.R. and M. Dowsett, *Aromatase and Its Inhibitors: Significance for Breast Cancer Therapy*. Recent progress in hormone research, 2002. **57**: p. 22.
-

40. Bulun, S.E., et al., *Aromatase excess in cancers of the breast, endometrium and ovary*. Journal of Steroid Biochemistry and Molecular Biology, 2007. **106**(1-5): p. 27.
41. Larionov, A., L. Berstein, and W. Miller, *Local uptake and synthesis of oestrone in normal and malignant postmenopausal breast tissues*. Journal of Steroid Biochemistry and Molecular Biology, 2002. **81**(1): p. 8.
42. Santen, R.J., *Inhibition of aromatase: insights from recent studies*. Steroids, 2003. **68**: p. 9.
43. Healey, C.S., et al., *Polymorphisms in the human aromatase cytochrome P450 gene (CYP19) and breast cancer risk*. Carcinogenesis, 2000. **21**(2): p. 5.
44. Robert, J., et al., *Predicting drug response and toxicity based on gene polymorphisms*. Critical Reviews in Oncology/Hematology, 2005. **54**: p. 26.
45. Dunning, A.M., et al., *Polymorphisms Associated With Circulating Sex Hormone Levels in Postmenopausal Women*. Journal of the National Cancer Institute, 2004. **96**(12): p. 10.
46. Akhtar, M., et al., *The impact of aromatase mechanism on other P450s*. Journal of Steroid Biochemistry and Molecular Biology, 1997. **61**(3-6): p. 6.
47. Corbin, C., J. Trant, and A. Conley, *Porcine gonadal and placental isozymes of aromatase cytochrome P450: sub-cellular distribution and support by NADPH-cytochrome P450 reductase*. Molecular and Cellular Endocrinology, 2000. **172**: p. 10.
48. Santen, R.J., et al., *History of Aromatase: Saga of an Important Biological Mediator and Therapeutic Target*. Endocrine Reviews, 2009. **30**: p. 34.
49. Simpson, E.R., et al., *Expression of the CYP19 (aromatase) gene: an unusual case of alternative promoter usage*. The FASEB Journal, 1997. **11**: p. 8.
50. Grodin, J., P. Siiteri, and P. MacDonald, *Source of estrogen production in postmenopausal women*. Journal of Clinical Endocrinology & Metabolism 1973 **36**(2): p. 8.
51. Roselli, C., et al., *Sexual partner preference, hypothalamic morphology and aromatase in rams*. Physiological Behavior, 2004. **83**: p. 13.
52. Boon, W., et al., *Hippocampal NMDA receptor subunit expression and watermaze learning in estrogen deficient female mice*. Brain Research - Molecular Brain Research, 2005. **140**: p. 6.
53. Resnik, R., et al., *The stimulation of uterine blood flow by various estrogens*. Endocrinology, 1974. **94**(4): p. 5.

-
54. Gennari, L., R. Nuti, and J. Bilezikian, *Aromatase activity and bone homeostasis in men*. Journal of Clinical Endocrinology & Metabolism, 2004. **89**: p. 10.
 55. Jones, M., et al., *Recognizing rare disorders: aromatase deficiency*. National Clinical Practice of Endocrinology & Metabolism, 2007. **3**: p. 8.
 56. Robertson, K., et al., *Impairment of spermatogenesis in mice lacking a functional aromatase (Cyp 19) gene*. Proceedings of the National Academy of Science USA, 1999. **96**: p. 6.
 57. Ellem, S., et al., *Local aromatase expression in human prostate is altered in malignancy*. Journal of Clinical Endocrinology & Metabolism, 2004. **89**: p. 8.
 58. Bulun, S., et al., *Endocrine disorders associated with inappropriately high aromatase expression*. Journal of Steroid Biochemistry and Molecular Biology, 1997. **61**(3-6): p. 7.
 59. Mackay, A., et al., *Molecular response to aromatase inhibitor treatment in primary breast cancer*. Breast Cancer Research, 2007. **9**(3): p. 14.
 60. Girault, I., et al., *Real-time reverse transcription PCR assay of CYP19 expression: application to a well-defined series of post-menopausal breast carcinomas*. Journal of Steroid Biochemistry & Molecular Biology, 2002. **82**: p. 10.
 61. Price, T., et al., *Determination of aromatase cytochrome P450 messenger ribonucleic acid in human breast tissue by competitive polymerase chain reaction amplification*. Journal of Clinical Endocrinology & Metabolism, 1992. **74**(6): p. 6.
 62. Ackerman, G., et al., *Aromatization of androstenedione by human adipose tissue stromal cells in monolayer culture*. Journal of Clinical Endocrinology & Metabolism, 1981. **53**(2): p. 6.
 63. Meng, L., et al., *Tumor necrosis factor alpha and interleukin 11 secreted by malignant breast epithelial cells inhibit adipocyte differentiation by selectively down-regulating CCAAT/enhancer binding protein alpha and peroxisome proliferator-activated receptor gamma: mechanism of desmoplastic reaction*. Cancer Research, 2001. **61**(5): p. 6.
 64. Sasano, H., et al., *Immunolocalization of aromatase and other steroidogenic enzymes in human breast disorders*. Human Pathology, 1994. **25**(5): p. 6.
 65. Pauley, R., et al., *Regulation of CYP19 aromatase transcription in breast stromal fibroblasts*. Journal of Clinical Endocrinology & Metabolism, 2000. **85**: p. 10.
-

66. Harada, N., T. Utsumi, and Y. Takagi, *Molecular and epidemiological analyses of abnormal expression of aromatase in breast cancer*. Pharmacogenetics, 1995. **5**: p. 6.
67. Gradishar, W.J., *Tamoxifen - What Next?* The Oncologist, 2004. **9**: p. 7.
68. Buzdar, A., *Anastrozole (Arimidex™) – an aromatase inhibitor for the adjuvant setting?* British Journal of Cancer, 2001. **85**(Suppl 2): p. 5.
69. Lake, D.E. and C. Hudis, *Aromatase Inhibitors in Breast Cancer: An Update*. Cancer Control, 2002. **9**(6): p. 9.
70. Kiang, D., et al., *Comparison of tamoxifen and hypophysectomy in breast cancer treatment*. Cancer, 1980. **45**(6): p. 4.
71. Strasser-Weippl, K. and P.E. Goss, *Advances in Adjuvant Hormonal Therapy for Postmenopausal Women*. Journal of Clinical Oncology, 2005. **23**(8): p. 9.
72. Gonnelli, S. and R. Petrioli, *Aromatase inhibitors, efficacy and metabolic risk in the treatment of postmenopausal women with early breast cancer*. Clinical Interventions in Aging, 2008. **3**(4): p. 12.
73. Buzdar, A.U., *Advances in Endocrine Treatments for Postmenopausal Women with Metastatic and Early Breast Cancer*. The Oncologist, 2003. **8**: p. 7.
74. Hu, J.C. and K. Mokbel, *Does c-erbB2/HER2 overexpression predict adjuvant tamoxifen failure in patients with early breast cancer?* European Journal of Surgical Oncology, 2001. **4**: p. 3.
75. Chen, S., et al., *What do we know about the mechanisms of aromatase inhibitor resistance?* Journal of Steroid Biochemistry and Molecular Biology, 2006. **102**(1-5): p. 18.
76. Dunbier, A.K., et al., *Progress in aromatase research and identification of key future directions*. Journal of Steroid Biochemistry & Molecular Biology, 2010. **118**: p. 5.
77. Goss, P., et al., *A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer*. New England Journal of Medicine, 2003. **349**: p. 10.
78. Khan, Q.J., A.P. O'Dea, and P. Sharma, *Musculoskeletal Adverse Events Associated with Adjuvant Aromatase Inhibitors*. Journal of Oncology, 2010. **2010**: p. 8.
79. Wright, C., et al., *Relationship between c-erbB-2 protein product expression and response to endocrine therapy in advanced breast cancer*. British Journal of Cancer, 1992. **65**(1): p. 4.
80. Ellis, M., et al., *Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive*

- primary breast cancer: evidence from a phase III randomized trial*. Journal of Clinical Oncology, 2001. **19**: p. 14.
81. Marcom, P., et al., *The combination of letrozole and trastuzumab as first or secondline biological therapy produces durable responses in a subset of HER2 positive and ER positive advanced breast cancers*. Breast Cancer Research and Treatment, 2007. **102**: p. 7.
82. Bardou, V., et al., *Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases*. Journal of Clinical Oncology, 2003. **21**: p. 7.
83. Dowsett, M., *Analysis of time to recurrence in the ATAC (Arimidex, tamoxifen, alone or in combination) trial according to estrogen receptor and progesterone receptor status*. Breast Cancer Research and Treatment, 2003. **82**(S7).
84. Brodie, A., G. Sabnis, and L. Macedo, *Xenograft models for aromatase inhibitor studies*. Journal of Steroid Biochemistry and Molecular Biology, 2007. **106**(1-5): p. 6.
85. Massarweh, S. and R. Schiff, *Unraveling the mechanisms of endocrine resistance in breast cancer: new therapeutic opportunities*. Clinical Cancer Research, 2007. **13**: p. 5.
86. Evans, W.E. and H.L. McLeod, *Pharmacogenomics — Drug Disposition, Drug Targets, and Side Effects*. The new england journal of medicine, 2003. **348**(6): p. 12.
87. Ma, C.X., et al., *Human Aromatase: Gene Resequencing and Functional Genomics*. Cancer Research, 2005. **65**(23): p. 12.
88. Gabriel, S., et al., *The structure of haplotype blocks in the human genome*. Science, 2002. **296**: p. 5.
89. Dunning, A.M., et al., *A Systematic Review Of Genetic Polymorphisms and Breast Cancer Risk*. Cancer Epidemiology, Biomarkers & Prevention, 1999. **8**: p. 12.
90. Talbott, K.E., et al., *A CYP19 (aromatase) polymorphism is associated with increased premenopausal breast cancer risk*. Breast Cancer Research and Treatment, 2008. **111**: p. 7.
91. Long, J.-R., et al., *Genetic Polymorphisms of the CYP19A1 Gene and Breast Cancer Survival*. Cancer Epidemiology, Biomarkers & Prevention, 2006. **15**(11): p. 8.

92. Kristensen, V.N., et al., *Genetic variants of CYP19 (aromatase) and breast cancer risk*. *Oncogene*, 2000. **19**(19): p. 5.
93. Haiman, C.A., et al., *A tetranucleotide repeat polymorphism in CYP19 and breast cancer risk*. *International Journal of Cancer*, 2000. **87**: p. 7.
94. Wang, L., et al., *Functional Genetic Polymorphisms in the Aromatase Gene CYP19 Vary the Response of Breast Cancer Patients to Neoadjuvant Therapy with Aromatase Inhibitors*. *Cancer Research*, 2010. **70**(1): p. 10.
95. Goode, E.L., et al., *Effect of Germ-Line Genetic Variation on Breast Cancer Survival in a Population-based Study*. *CANCER RESEARCH*, 2002. **62**: p. 6.
96. Garcia-Casado, Z., et al., *A polymorphism at the 3'-UTR region of the aromatase gene defines a subgroup of postmenopausal breast cancer patients with poor response to neoadjuvant letrozole*. *BMC Cancer*, 2010. **10**(36): p. 11.
97. Baxter, S.W., et al., *Polymorphic variation in CYP19 and the risk of breast cancer*. *Carcinogenesis*, 2001. **22**(2): p. 3.
98. Susptsin, E.N., et al., *Distinct prevalence of the CYP19 3(TTTA)7 allele in premenopausal versus postmenopausal breast cancer patients, but not in control individuals*. *European Journal of Cancer*, 2002. **38**: p. 6.
99. Miyoshi, Y., et al., *Breast cancer risk associated with polymorphism in CYP19 in Japanese women*. *International Journal of Cancer*, 2000. **89**(4): p. 4.
100. Huang, C., et al., *The CYP19 TTTA Repeat Polymorphism Is Related to the Prognosis of Premenopausal Stage I-II and Operable Stage III Breast Cancers*. *The Oncologist*, 2008. **13**: p. 11.
101. Haiman, C.A., et al., *Genetic Variation at the CYP19A1 Locus Predicts Circulating Estrogen Levels but not Breast Cancer Risk in Postmenopausal Women*. *Cancer Research*, 2007. **67**(5): p. 5.
102. Haiman, C.A., et al., *No Association between a Single Nucleotide Polymorphism in CYP19 and Breast Cancer Risk*. *Cancer Epidemiology, Biomarkers & Prevention*, 2002. **11**: p. 2.
103. Cai, H., et al., *Association of genetic polymorphisms in CYP19A1 and blood levels of sex hormones among postmenopausal Chinese women*. *Pharmacogenetic Genomics*, 2008. **18**(8): p. 14.
104. Ralph, D.A., et al., *Age-specific Association of Steroid Hormone Pathway Gene Polymorphisms With Breast Cancer Risk*. *Cancer*, 2007. **109**(10): p. 9.
105. Colomer, R., et al., *A Single-Nucleotide Polymorphism in the Aromatase Gene Is Associated with the Efficacy of the Aromatase Inhibitor Letrozole in Advanced Breast Carcinoma*. *Clinical Cancer Research*, 2008. **14**(3): p. 6.

106. Guo, Y., et al., *Polymorphisms of estrogen-biosynthesis genes CYP17 and CYP19 may influence age at menarche: a genetic association study in Caucasian females*. Human Molecular Genetics, 2006. **15**(16): p. 14.
107. Xita, N., et al., *The (TTTA)_n polymorphism of aromatase (CYP19) gene is associated with age at menarche*. Human Reproduction, 2010. **25**(12): p. 5.
108. Park, I.H., et al., *Single nucleotide polymorphisms of CYP19A1 predict clinical outcomes and adverse events associated with letrozole in patients with metastatic breast cancer*. Cancer Chemotherapy and Pharmacology, 2011: p. 9.

