Università degli Studi di Padova

Dipartimento di Scienze del Farmaco

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE FARMACOLOGICHE

INDIRIZZO FARMACOLOGIA MOLECOLARE E CELLULARE

CICLO XVII

INNOVATIVE STRATEGIES FOR TAILORING THERAPY IN CANCER PATIENT:

PHARMACOGENETICS AND HORMONE THERAPY PERSONALIZATION IN METASTATIC OR LOCALLY ADVANCED BREAST CANCER PATIENTS TREATED WITH EXEMESTANE

Direttore della Scuola: Ch.mo Prof. Pietro Giusti

Coordinatore d'indirizzo: Ch.mo Prof. Pietro Giusti

Supervisore: Ch.mo Prof. Pietro Giusti

Supervisore esterno: Dott. Giuseppe Toffoli

Dottoranda: Sara Gagno

Alla mia famiglia

Questo lavoro di tesi è stato svolto presso la scuola di dottorato in Scienze Farmacologiche, indirizzo Farmacologia Cellulare e Molecolare dell'Università degli Studi di Padova, diretta dal Prof. Pietro Giusti in collaborazione con la Struttura Operativa Complessa di Farmacologia Sperimentale e Clinica del Centro di Riferimento Oncologico di Aviano (Istituto di ricerca e cura a carattere scientifico), diretta dal Dott. Giuseppe Toffoli.

CONTENTS

ABSTRACT

Background: Response to chemotherapeutic agents is highly variable among patients both in terms of efficacy and tolerability; consequently personalization of drug therapy is one of the main objective in cancer treatment in order to reduce adverse drug reactions (ADRs), improve efficacy while decreasing the costs of treatments. Many factors account for inter-individual differences. Among them, patient's genetic background has attracted interest for personalization of drug therapy (Pharmacogenetics).

Breast cancer (BC) is the female most frequently diagnosed malignancy and the primary cause of cancer-related death among females with 1.380.000 new cases and 458.000 deaths worldwide registered in 2008. Estrogen receptors (ER) are over-expressed in around 80% of BC cases and ER-positive (ER+) cancer cells depend on estrogens for their growth. In postmenopausal women estrogens can derive only from androgens through an aromatization reaction. Aromatase (CYP19A1) is a key enzyme in this process and, for this reason, is the target of many inhibitors drugs, including exemestane. Anti-aromatase treatments represent the current corner-stone of ER+ BC therapy in postmenopausal women.

Exemestane is a steroidal irreversible third generation aromatase inhibitor (AI) which determines the inactivation of the enzyme, resulting in estrogen synthesis inhibition and deprivation.

Exemestane is used in adjuvant setting for ER+ early stage invasive BC and for the treatment of advanced stage BC progressed to a previous anti-estrogen therapy.

Several germ line variations (polymorphisms) have been described in genes involved either in estrogens activity and metabolism or in the pharmacokinetics of exemestane.

Aims: this PhD thesis had a dual aim:

- setting up a pharmacogenetic method to analyze estrogen and exemestane-related polymorphisms (Single Nucleotide Polymorphisms (SNPs) and Short Tandem Repeats (STRs)),
- determining the predictive and prognostic value of these polymorphisms in postmenopausal metastatic or locally advanced ER+ BC patients (Response Rate (RR), Clinical Benefit (CB), Time To Progression (TTP) and Overall Survival (OS)).

Fifteen polymorphisms in genes involved estrogens synthesis (*CYP17A1* and *CYP19A1*), activity (*ESR1*, *ESR2* and *RIZ1*) and metabolism (*CYP1B1*, *UGT1A1* and *COMT*) as well as genes implicated in the metabolic pathway of exemestane (*CYP3A4* and *CYP3A5*) were investigated.

We considered the CYP19A1_Ex11_410A/C (rs4646) SNP in the 3' untraslated region (3'UTR) of the aromatase gene, previously associated to a better disease-free survival (DFS) (Colomer *et al.*, 2008) in patients treated with the AI, letrozole, and to a better OS in patients treated with another AI,

anastrozole (Liu *et al.*, 2013). However, the role of this polymorphism has not yet been clearly defined. Other *CYP19A1* polymorphisms analyzed were: CYP19A1_47T/C (rs700519), CYP19A1_1558T/C (rs10046) and CYP19A1_(TTTA)n (rs60271534), along with a SNP on the *CYP17A1* gene (CYP17A1_27A/G (rs743572)), coding for another enzyme responsible for estrogens synthesis.

Additionally, we investigated:

- polymorphisms on estrogens receptors: *ESR1*: ESR1_497T/C (rs2234693), and ESR1_256A/G (rs9340799); *ESR2:* ESR2_1082A/G (rs1256049), and ESR2_1730A/G (rs4986938); *RIZ1*: RIZ1_delP704 (rs2308040);
- polymorphisms on estrogens metabolizing enzymes: *CYP1B1*: CYP1B1*3_4326G/C (rs1056836), *UGT1A1*: UGT1A1*28_ TA(6/7) (rs8175347) and *COMT*: COMT_12A/G (rs4680);
- polymorphisms on enzymes involved in the oxidative metabolism of exemestane: *CYP3A4*: CYP3A4*1B_-392A/G (rs2740574) and *CYP3A5*: CYP3A5*3_6986A/G (rs776746).

Methods: genetic analyses were conducted in a group of 275 ER+ metastatic or locally advanced BC patients treated with exemestane as first line hormone therapy.

Patients were subjected to blood sampling before the beginning of therapy. DNA was extracted from whole blood, and then amplified by Polymerase Chain Reaction (PCR).

Four methods for polymorphisms genotyping were set up and developed: Pyrosequencing, TaqMan® Allelic Discrimination Assay, Automated Fragment Analysis and Illumina GoldenGate Assay.

Statistical associations between genetic determinants and clinical outcome were assessed by the two-sided Fisher's Exact Test (associations between genotypes and clinical responses) and the Kaplan-Meier product-limit method with the log-rank test statistic (associations between polymorphisms and TTP/OS).

Results: For each polymorphism the most appropriate technique, based on the best result obtained in the setting up process, was chosen. As a result:

- three SNPs were investigated with Pyrosequencing: CYP19A1_47T/C (rs700519), CYP3A4*1B_-392A/G (rs2740574), and RIZ1_delP704 (rs2308040);
- ten SNPs were genotyped with TaqMan® Allelic Discrimination Assay: CYP19A1_Ex11+410A/C (rs4646), CYP19A1_1558T/C (rs10046), CYP3A5*3_6986A/G (rs776746), COMT_12A/G (rs4680), ESR1_497T/C (rs2234693), ESR1_256A/G (rs9340799), ESR2_1082A/G (rs1256049), ESR2_1730A/G (rs4986938), CYP17A1_27A/G (rs743572) and CYP1B1*3_4326G/C (rs1056836);
- two STR were analyzed with Automated Fragment Analysis: CYP19A1_(TTTA)n (rs60271534) and UGT1A1*28_ TA(6/7) (rs8175347)

- twelve of the above mentioned polymorphisms were additionally analyzed by Illumina GoldenGate Assay as positive controls. The result obtained by this validation process was a 100% accordance within the obtained genotypes.

Among the polymorphisms investigated, a statistically significant association was observed for *CYP1B1*, the gene encoding for the enzyme which catalyze the phase I estrogens oxidative metabolism. The variant (G) allele of CYP1B1*3_4326G/C (rs1056836) was significantly associated with clinical response to exemestane (RR, OR_{GG} = 2.91, 95% CI = 5.88 – 1.25, p = 0.0039; according to the two-sided Fisher's exact test). The same variant allele was also significantly associated with the TTP and OS (TTP, dominant model: HR_{CG+GG}= 0.66, 95% CI = 0.50 – 0.87, p = 0.0037; OS, dominant model = HR_{CG+GG}= 0.66, 95% CI = 0.46 – 0.95, p = 0.023, according to the log-rank test) meaning that patients carrying at least one variant allele (G) not only showed a better clinical response, but experienced also a later progression and a longer survival than wild type patients.

Regarding the aromatase gene (*CYP19A1* gene), the only association found, even if marginal, was between CYP19A1_1558T/C (rs10046) SNP and TTP (HR_{CC} recessive model=1.4, 95% CI = 1.04 – 1.89, p = 0.028). Conversely, we did not find any significant association between CYP19A1_Ex11_410A/C (rs4646) SNP, (the main objective of the study) and RR, CB, TTP or OS, respectively.

Concerning aromatase gene polymorphisms, we were able to describe a new genetic variant for the CYP19A1_(TTTA)n (rs60271534) STR in intron 4. Genetic databases and literature report that the number of repeats varies from 7 to 13, but we found a still not described 14 (TTTA) repeats allele.

Conclusions: in conclusion, this thesis work allowed defining a new molecular marker, CYP1B1*3_4326G/C (rs1056836) SNP, with a predictive and prognostic value for exemestane-based treatment of postmenopausal ER+ metastatic or locally advanced BC patients. This indicates that, once validated, this marker could potentially be employable in the daily clinical oncology practice as a tool which may allow the identification of patients more likely to be responsive to treatment by a simple genetic evaluation from peripheral blood, performed prior to therapy. In addition, we described a new genetic variant in the aromatase gene.

RIASSUNTO

Introduzione: La risposta agli agenti chemioterapici è altamente variabile tra i pazienti sia per quanto riguarda l'efficacia che la tollerabilità, di conseguenza la personalizzazione della terapia è uno dei principali obiettivi della ricerca in campo oncologico con l'obiettivo di ridurre le reazioni avverse al farmaco, migliorarne l'efficacia e nel contempo contenerne i costi.

I fattori responsabili della variabilità interindividuale sono molteplici. Tra questi, il *background* genetico dei pazienti ha attratto interesse per la personalizzazione della terapia (Farmacogenetica).

Il carcinoma mammario (*breast cancer* - BC) rappresenta la neoplasia più frequentemente diagnosticata e la prima causa di morte collegata al cancro tra le donne. Nel 2008 sono stati registrati, a livello mondiale, 1.380.000 nuovi casi e 458.000 morti a causa del cancro della mammella.

Il recettore degli estrogeni (*estrogen receptor* - ER) risulta iper-espresso in circa l'80% dei casi di BC e le cellule cancerose positive al ER (ER+) dipendono dagli estrogeni per la loro crescita. Nelle donne in menopausa, gli estrogeni derivano unicamente dagli androgeni attraverso una reazione di aromatizzazione. L'aromatasi (CYP19A1) è un enzima chiave in questo processo e, per questa ragione, è diventato il target di numerosi farmaci inibitori, compreso exemestane. Il trattamento anti-aromatasi rappresenta attualmente il cardine della terapia del ER+ BC nelle donne in menopausa.

Exemestane è un inibitore irreversibile dell'aromatasi (AI) di terza generazione e di tipo steroideo che determina l'inattivazione dell'enzima, provocando quindi l'inibizione della sintesi estrogenica.

Exemestane è un farmaco impiegato in assetto adiuvante per il ER+ BC invasivo allo stadio precoce ed in assetto avanzato se la malattia è progredita dopo una precedente terapia anti-estrogenica.

Sono state descritte numerose variazioni genetiche germinali (polimorfismi) in geni coinvolti sia nell'attività e metabolismo degli estrogeni che nella farmacocinetica di exemestane.

Obiettivi: questa tesi di dottorato ha avuto un duplice obiettivo:

- mettere a punto un metodo di indagine farmacogenetica per analizzare polimorfismi correlati ad estrogeni ed exemestane (polimorfismi a singolo nucleotide – SNPs e microsatelliti (*short tandem repeats*) - STRs)
- determinare il ruolo predittivo e prognostico di tali polimorfismi come biomarcatori di efficacia del trattamento a base di exemestane, in termini di *Response Rate* (RR), *Clinical Benefit* (CB), tempo alla progressione (TTP) e sopravvivenza globale (OS).

Sono stati considerati quindici polimorfismi in geni coinvolti nella sintesi (*CYP17A1* e *CYP19A1*), attività (*ESR1*, *ESR2* e *RIZ1*) e metabolismo (*CYP1B1*, *UGT1A1* e *COMT*) degli estrogeni insieme a geni implicati nel *pathway* metabolico di exemestane (*CYP3A4* e *CYP3A5*).

Come obiettivo primario dello studio clinico è stato considerato lo SNP CYP19A1_Ex11_410A/C (rs4646) della regione non 3' tradotta (*3' untraslated region* - 3'UTR) del gene dell'aromatasi, già in precedenza associato ad una migliore sopravvivenza libera da malattia (*disease free servival* – DFS) (Colomer *et al.*, 2008) in pazienti trattate con l'AI letrozolo e con la miglior OS in pazienti trattate con un altro AI, l'anastrozolo (Liu *et al.*, 2013). Ciononostante, il ruolo di questo polimorfismo non è stato ancora chiaramente definito. Sono stati analizzati anche altri polimorfismi del gene *CYP19A1* (CYP19A1_47T/C (rs700519), CYP19A1_1558T/C (rs10046) e CYP19A1_(TTTA)n (rs60271534)) insieme ad uno SNP nel gene *CYP17A1*(CYP17A1_27A/G (rs743572)), codificante per un altro enzima responsabile della sintesi degli estrogeni.

Inoltre, sono stati indagati polimorfismi dei geni codificanti per:

- i recettori degli estrogeni: *ESR1*: ESR1_497T/C (rs2234693), e ESR1_256A/G (rs9340799); *ESR2:* ESR2_1082A/G (rs1256049), e ESR2_1730A/G (rs4986938); *RIZ1*: RIZ1_delP704 (rs2308040);
- gli enzimi deputati al metabolismo degli estrogeni: *CYP1B1*: CYP1B1*3_4326G/C (rs1056836), *UGT1A1*: UGT1A1*28_ TA(6/7) (rs8175347) e *COMT*: COMT_12A/G (rs4680);
- gli enzimi responsabili del metabolismo ossidativo di exemestane: *CYP3A4*: CYP3A4*1B_-392A/G (rs2740574) e *CYP3A5*: CYP3A5*3_6986A/G (rs776746).

Metodi: le analisi genetiche sono state condotte in un gruppo di 275 pazienti affetti da ER+ BC metastatico o localmente avanzato trattate con exemestane come prima linea di trattamento ormonale. Ai pazienti è stato effettuato un prelievo ematico prima dell'inizio della terapia. Il DNA è stato poi estratto dal campione di sangue intero ed amplificato tramite la reazione a catena della polimerasi (PCR).

Per le analisi genetiche sono state messe a punto quattro tecniche di genotipizzazione: Pyrosequencing, Saggio di Discriminazione Allelica mediante sonde TaqMan®, Analisi dei Frammenti Automatizzata ed il saggio GoldenGate di Illumina.

Sono state valutate le associazioni statistiche tra i determinanti genetici e l'*outcome* clinico dei pazienti attraverso il Test Esatto di Fisher a due vie per l'associazione di polimorfismi e risposta clinica e attraverso lo stimatore del prodotto limite di Kaplan Meier e il test dei ranghi logaritmici per l'associazione tra polimorfismi e TTP/OS.

Risultati: per ogni polimorfismo è stata scelta la tecnica di indagine molecolare più appropriata a seconda del miglior risultato ottenuto durante la fase di messa a punto delle metodologie. Di conseguenza:

- tre SNPs sono stati analizzati con il Pyrosequencing: CYP19A1_47T/C (rs700519), CYP3A4*1B_-392A/G (rs2740574), e RIZ1_delP704 (rs2308040);

- dieci SNPs sono stati genotipizzati con il saggio di Discriminazione Allelica mediante sonde TaqMan®: CYP19A1_Ex11+410A/C (rs4646), CYP19A1_1558T/C (rs10046), CYP3A5*3_6986A/G (rs776746), COMT_12A/G (rs4680), ESR1_497T/C (rs2234693), ESR1_256A/G (rs9340799), ESR2_1082A/G (rs1256049), ESR2_1730A/G (rs4986938), CYP17A1_27A/G (rs743572) e CYP1B1*3_4326G/C (rs1056836);
- due STR sono stati esaminati attraverso l'Analisi dei Frammenti Automatizzata: CYP19A1_(TTTA)n (rs60271534) e UGT1A1*28_ TA(6/7) (rs8175347);
- i campioni analizzati per dodici dei sopraccitati polimorfismi sono stati, inoltre, inclusi nel saggio Illumina GoldenGate come controlli positivi. Il risultato di questo processo di validazione è stata una concordanza del 100% tra i genotipi ottenuti con questa tecnica e quelli derivanti dalle precedenti indagini.

Tra i polimorfismi analizzati, è stata osservata un'associazione statisticamente significativa per *CYP1B1*, gene codificante per l'enzima responsabile del metabolismo ossidativo di prima fase degli estrogeni. L'allele variante G del polimorfismo CYP1B1*3_4326G/C (rs1056836) è stato significativamente associato con la risposta clinica ad exemestane (RR, OR_{GG} = 2.91, 95% CI = 5.88 – 1.25, p = 0.0039; secondo il Test Esatto di Fisher a due vie). Lo stesso allele variante è stato significativamente associato anche al TTP e alla OS (TTP, modello dominante: HR_{CG+GG}= 0.66, 95% CI = 0.50 – 0.87, p = 0.0037; OS, modello dominante = HR_{CG+GG}= 0.66, 95% CI = 0.46 – 0.95, p = 0.023, secondo il test dei ranghi logaritmici). Questo significa che pazienti portatori di almeno un allele G non solo hanno dimostrato una miglior risposta clinica al trattamento ma hanno anche avuto una progressione più tardiva ed una sopravvivenza più lunga dei pazienti *wild type*.

Per quanto riguarda il gene dell'aromatasi, l'unica associazione riscontrata, anche se marginale, riguarda il polimorfismo CYP19A1_1558T/C (rs10046) il cui allele variante C che è stato associato ad un ridotto TTP (HR_{CC} modello recessivo =1.4, 95% CI = 1.04 – 1.89, p = 0.028, secondo il Test Esatto di Fischer a due vie).

Al contrario, non è stata riscontrata alcuna associazione significativa tra lo SNP CYP19A1_Ex11_410A/C (rs4646), obiettivo principale dello studio, e RR, CB, TTP o OS.

Riguardo i polimorfismi del gene dell'aromatasi, siamo stati in grado di descrivere una nuova variante genetica per il polimorfismo STR CYP19A1_(TTTA)n (rs60271534) dell'introne 4. Le banche dati genetiche e la letteratura riportano che il numero di ripetizioni della quadripletta TTTA vari tra 7 e 13, ma nel nostro studio è stato individuato un allele, finora mai descritto, con 14 ripetizioni.

Conclusioni: in conclusione, questo lavoro di tesi ha permesso di definire un nuovo biomarcatore molecolare, lo SNP CYP1B1*3_4326G/C (rs1056836), con un valore predittivo e prognostico per il trattamento a base di exemestane in pazienti affetti da ER+ BC, metastatico o localmente avanzato.

Questo presuppone che, se validato, questo biomarcatore potrebbe potenzialmente essere impiegato nella pratica clinica oncologica quotidiana come strumento che potrebbe aiutare ad identificare i pazienti che hanno una maggiore probabilità di risposta all'exemestane tramite una semplice valutazione genetica da sangue periferico da effettuarsi prima della terapia. Inoltre, è stata descritta una nuova variante genetica del gene dell'aromatasi.

1.INTRODUCTION

1.1 PHARMACOGENETICS

Pharmacogenetics is the branch of pharmacology which aims at studying the role of inherited genetic differences in affecting individual responses to drugs, both in terms of therapeutic and adverse effects.
One of the major issues in oncology is that the response to anticancer drugs is highly variable within individuals: as reviewed by Robert and colleagues (Robert *et al.,* 2014), the best protocols, applied to sensitive cancers barely provide not more than 50% responses. The exceptions are few and this means that approximately 50% of patients receive an often toxic and costly treatment without any beneficial effect.
There are several examples in literature reporting that the administration of the same dose of the same antineoplastic drug often shows a huge range of responses and toxicities (which in some cases can result even lethal) among patients affected by the same cancer disease.
Moreover, anticancer agents have a low therapeutic index (i.e. the ratio of the highest exposure to the drug that results in no toxicity to the exposure that produces the desired efficacy (Muller *et al.*, 2012)): this leads to a high risk of developing adverse effects in some patients due to genetic and environmental reasons. Several host-related factors could be responsible for this variability in drug response and toxicities: age, sex, hepatic and renal functions, comorbidities and co-medications, etc.
Inter-individual differences in pharmacokinetics (PK), that is, drug absorption, distribution, metabolism, and elimination, and in pharmacodynamics (PD), i.e., effects on drug receptors and other drug targets lead as well to an inter-patients different behavior towards the drug.
In addition, variations in the structure or in the expression of genes that encode proteins involved in the PK and PD of a drug (target enzymes or proteins related to its function) could significantly contribute to individual differences in drug response.
Genetic variability consists of several biological mechanisms: there are differences in transcription factor activity, gene expression, gene silencing (epigenetics), and genetic polymorphism.
Current clinical practice bases the prescription of anticancer drugs as a function of patients' features.
There are numerous prognostic factors in which oncologist can rely on, derived from clinical data (tumour size, node invasion, performance status) or from pathological data (histoprognostic grading). Nowadays, the dose of most anticancer drugs is generally based on the individual's body weight and body surface area. However, this strategy, with undoubted value to guide therapeutic decisions, is often not enough to overcome the inter-individual differences observed in the outcome of treatment (Efferth *et al.*, 2005; Marsh *et al.*, 2006) and are not predictive or not helpful for the choice of the best drug or the best drug combination, in order to achieve the highest efficacy and the minimal toxicity.
Pharmacogenetics is a promising tool for overcoming patients' variability. Pharmacogenetics intends to identify relationships between gene polymorphisms and drug activity, in order to propose a rational and

tailored drug prescription by the prediction of drug efficacy and toxicity.

When referring to genetic influence on the treatment outcome there is still some confusion between the terms “pharmacogenetics” and “pharmacogenomics”. Indeed, they tend to be used interchangeably and a precise consensus definition of either term remains equivocal. On the one hand, “pharmacogenetics” focuses on the association of candidate genes polymorphisms with drug activity, while “pharmacogenomics” considers the entire genome, through the broader application of new genomic technologies. On the other hand, in oncology “pharmacogenetics” is often considered as concerning the germ-line polymorphisms and individual patient’s features, while “pharmacogenomics” usually refers to those of the tumor.

1.1.1 Polymorphisms

The great majority of human DNA sequence is identical among individuals, except for minor changes called polymorphisms constituted by nucleotide substitutions, deletions and insertions, repeats, gene copy number variations and sometimes more important rearrangements. Structure, expression, stability and activity of the proteins encoded by genes can be affected by DNA polymorphisms. Leading to minor phenotypic variations, polymorphisms explain the inter-individual differences, from eyes or hair color to disease susceptibility or drug sensitivity: this is why they present a major interest by a clinical point of view.

Substitution, deletion or insertion of nucleotides can arise from errors occurring during DNA replication or lesions induced by mutagenic agents which may lead to the replacement of a nucleotide by another one (substitution), to the loss (deletion) or to the addition of a nucleotide (insertion). When the coding sequence of a gene is affected by polymorphisms, the encoded protein may bear structural alterations, which possibly lead to its instability, an enhancement or reduction in its activity or the loss of its functionality.

“Single Nucleotide Polymorphisms” or SNPs are one of the most common forms of genetic variations (>90%) and are characterized by the involvement of only one nucleotide (by substitution, insertion or deletion).

Some polymorphisms are instead more complex: they concern the number of short series of nucleotides repeats (minisatellite or microsatellite), called Variable Number of Tandem Repeats (VNTR), or even the number of copies of a gene (CNV, Copy Number Variations). These polymorphisms are more likely to play a major role in the level of the mRNA and the protein produced.

Conventionally, a SNP is defined as a nucleotide variation having an allele frequency greater than 1%, whereas, when the frequency is lower, the genetic variation is indicated as mutation. Mutations by substitution even if are biochemically identical to SNPs (both involving the replacement of a nucleotide by another one), have a different meaning: they are rare and deleterious while SNPs are common and

non-deleterious events.

The frequency of gene polymorphisms is generally higher in introns than in exons.

Polymorphic genes have usually not indispensable functions for cell life: polymorphisms in genes involved in vital processes, if deleterious, would be rapidly eliminated by natural selection. In contrast, genes coding for enzymes involved in the metabolism of xenobiotics and drugs are often polymorphic since they have no major consequences on cell viability and so they are not eliminated from the genome. Nevertheless, an effect of these polymorphisms may arise in some special situations such as contact with xenobiotics or DNA damaging agents in which a polymorphic variant could allow some flexibility in front of environmental variations.

Heterogeneous systems are used for SNPs nomenclature but the universal accepted one is that using the rs code (reference sequence). This system is the only one allowing the precise identification of a polymorphic variation within the most common used genetic databases (NCBI, HapMap, SNP500 Cancer, 1000 genomes, etc.). If the SNP leads to the replacement of an amino acid by another one (non-synonymous SNP), usually the SNP nomenclature is completed by the name and the position of the amino-acid that is replaced in the protein, followed by the name of the novel amino acid (*i.e*, L432V or Leu432Val).

1.1.2 Types of polymorphisms

SNPs lying in the coding region can be classified as:

- *synonymous* or *silent* when there is no change of the amino acid encoded: since the genetic code is degenerated, distinct codons can encode the same amino acid. However, these polymorphisms may have functional consequences since the frequency of the various codons corresponding to the same amino acid is variable, and the corresponding tRNA population may not be adapted to the replacement of a codon by another one. Accordingly, if the tRNAs corresponding to the variant codon are rarer than the wild-type ones, the synthesis rate or the protein folding may be affected. Moreover, the replacement of a nucleotide by another one may produce a different three-dimensional structure of the mRNA or a different stability of the variant transcript and this may have consequences on the rate of its translation into protein, therefore on the amount of the protein synthesized.
- *missense* when there is a replacement of the amino-acid by another one. If the substituted amino-acid shares the same chemical properties of the wild-type one (e.g. valine replaced by leucine, glutamic acid replaced by aspartic acid) it usually produces a modest effect. In contrast, if an hydroxylated amino acid residue (potential substrate of a protein kinase) is replaced, for instance, by an aliphatic amino acid residue, an acidic by a basic one, functional effects may be more important.

- *non-sense* if the polymorphism imply the occurrence of a stop codon leading to a truncated protein. The same effect could be obtained if the alteration occurs within a splicing site.
- *frameshift polymorphism* when, as suggested by the definition, insertions and deletions alter the reading frame, therefore generating completely different codons then changing the entire downstream sequence of the protein. Usually leading to the generation of early stop codons, these variations have a high probability to induce the production of a truncated, totally inactive protein when affecting the coding sequence. Less deleterious effect is instead produced when they occur within the gene regulatory regions, especially in the 3' and 5' untranslated regions (UTR), or in intronic sequences, where they can lead to alterations in protein expression.

When SNPs occur at the splicing site an abnormal protein is produced, because the enzymes responsible for mRNA maturation no longer recognize the splicing site. The resulting protein could either lack the portion encoded by the missing exon or including an intronic aberrant sequence rapidly concluded by a stop codon. The SNP rs776746 of the CYP3A5 gene is an example of a splicing site polymorphism, leading to the complete absence of the protein in about 90% of Caucasian subjects (Xie *et al.*, 2004).

Many polymorphisms lie in non-coding regions like introns and 5' or 3' UTR with possible phenotypic effect if these sequences have regulatory functions (promoters, silencers, enhancers, micro-RNA binding sites or micro-RNA genes).

Repetitive sequences often consist of variable number repeats of the CA or TA dinucleotide within a microsatellite that may be involved in gene regulation. When the number of repeats ranges from 2 to 5 the polymorphism is called Short Tandem Repeats polymorphism (STR).

The TA repeat in the promoter of the UGT1A1 gene represents one the best known STR example: if 7 repeats instead of 6 are present there is a 50% decrease in gene transcription rate, with hyperbilirubinemia and increased risk of irinotecan toxicity as major phenotypic consequences (Innocenti *et al.*, 2003).

Finally, unequal chromosomal recombinations may lead to polymorphisms concerning an entire gene, producing a variable copy number of that gene. Usually, if the gene copy number increases, the encoded protein is overexpressed: this is the case, for instance, of the CYP2D6 gene, which is duplicated (or more) in about 5% of Caucasian subjects (Sachse *et al.*, 1997).

Gene deletions also exist: for instance, the glutathione transferase genes GSTM1 and GSTT1 are deleted in about 20% and 50% of Caucasian subjects, respectively (Kagimoto *et al.*, 1990).

1.1.3 Genotyping techniques

Genotyping is the process of determining the type of nucleotide that resides in a given polymorphic site by the use of molecular tools. It is performed by examining the individual's DNA sequence through biological assays and comparing it to a reference sequence. The increased interest in SNPs has been

reflected by the furious development of a diverse range of SNP genotyping methods, which allows low to high throughput genotyping.

Current methods of genotyping include (Table 1):

- targeted analyses, performed in order to investigate one or few specific SNPs at the same time, thus with a low throughput result. Techniques based on a targeted analyses are: restriction fragment length polymorphism (RFLP) identification, real-time based allelic discrimination (TaqMan® Assay), sequencing by synthesis (Pyrosequencing) technology, *de novo* DNA sequencing (conventional Sanger Sequencing);
- SNPs arrays: allowing the analysis of a wider panel of specific SNPs than the targeted analyses, SNPs arrays are considered medium-throughput technologies, for example bead arrays (GoldenGate, BeadXpress, xMAP® and xTAG®), mass spectrometry arrays (Sequenom), microarrays (DMET, GeneChip, Sentrix® Array Matrix, Sentrix® BeadChip), GWAS platforms (SNP 6.0, Axiom, OmniExpress, Omin1-Quad, Omni 5M)
- next genetration sequencing (NGS), performing a massive parallel sequencing is considered a high-throughput approach of DNA sequencing. NGS technologies includes, for example, 454 GS FLX+, HiSeq, MiSeq, Solid system.

The ability to determine genotypes of many individuals accurately and efficiently has allowed genetic studies that cover more of the variation within individual genes, instead of focusing only on one or a few coding variants, and to do so in study samples of reasonable power.

Method	Thechnology	Assay (Company)
RFLP	Enzymatic Digestion	Multiple
Real-time based	Allelic Discrimination	- TaqMan (AppliedBiosystem) - LightCycler (Roche)
Sequencing-by-synthesis	Pyrosequencing	- PSQ 96 (Qiagen)
De novo DNA sequencing	Conventional Sanger sequencing	Multiple
Bead array	BeadXpress, GoldenGate xMAP® and xTAG®	Illumina Luminex
Mass spectrometry based	MassARRAY system	Sequenom
Microarray platform	DMET platform, GeneChip platform Sentrix® Array Matrix, Sentrix® BeadChip	Affiyetrix Illumina
GWAS platform	SNP 6.0, Axiom OmniExpress,Omni1-Quad,Omni5M	Affiyetrix Illumina
Next generation Sequencing	454 GS FLX+ HiSeq System, MiSeq SOLID system	Roche Illumina AppliedBiosystem

Table 1. Conventional employed genotyping techniques. It is noteworthy that the throughput of these techniques goes from low to high as one moves down along the table.

1.1.4 Pharmacogenetic studies

There are three different approaches to design a pharmacogenetic study (Fig. 1):

1. *A candidate gene approach*:

   The name of this approach is due to the fact that some genes, considered of pharmacologic interest for the PK or PD of the drug, are chosen in advance as candidate genes before the study beginning. Therefore, a candidate gene approach is also termed hypothesis-driven association study. Typical candidate genes encode, for example, drug transporters, biotransformation enzymes, or drug receptors. In the candidate gene approach, only a reduced number of polymorphisms, selected on the basis of their function, are investigated. Usually, SNPs leading (or predicted to lead) to alterations in the protein function or expression profile are selected since they are likely to affect the drug response (Kooloos *et al*., 2009).

   This is a very reasonable approach; however, in many cases, one single genetic trait does not suitably explain the wide range of inter-individual differences in clinical outcome so the clear disadvantage in considering a single gene approach is the possibility of losing significant associations (Deenen *et al.*, 2011).

2. *A genome wide approach*: in contrast to the candidate gene approach, in which only a limited number of polymorphisms are tested, the genome-wide approach interrogates the entire human genome in multiple loci (mostly SNPs). Therefore, it is not a hypothesis-driven approach; on the contrary, it is independent of whether or not a gene is *a priori* expected to be involved in the pharmacological pathway of a drug. This approach requires high-throughput genotyping technologies that are able to analyze multiple SNPs simultaneously. This is allowed by the next-generation sequencing (NGS) technologies, which can investigate a number of SNPs that may range from a few hundred to even hundreds of thousands (Deenen *et al.*, 2011). There are three types of NGS: whole exome sequencing (which analyzes all the SNPs lying on the exons), whole genome sequencing (which interrogates SNPs on the entire genome), and targeted gene panels (which focuses on genes of interest SNPs). NGS have reduced the cost of large-scale sequencing by several orders of magnitude, and allow the identification of rare SNPs or mutations, which are lost with the previous approach. However, these technologies, take with them additional challenges involving the use of constantly evolving platforms, the clinical validation of the results obtained, and the management of the huge amount of data produced which requires bioinformatics expertise. (Rehm *et al.*, 2013). Moreover, the expression signals of irrelevant genes (defined as “noise”) that could increase the number of false positive and the high number of patients requested to have a statistical powered study are other disadvantages.

3. *A pathway gene approach*: as the name suggests, this method focuses on the genes of the entire drug pathway, combining the advantages of the single-gene and the genome wide

approaches. This approach takes into account two factors: from one hand, that the pharmacological pathway of a drug is very complex and involves many PK and PD proteins, from the other hand, that differences in response to anticancer drugs are mostly polygenetic traits. Hence, a set of SNPs in genes belonging to the pathway are selected and if some associations are found, they can be rationally attributed to alterations in one ore more steps of the drug's mechanism of action (Kooloos *et al*., 2009). The main advantage of the gene pathway approach is to combine information of several genes that share a common metabolic pathway and to minimize the "noise" of a non-targeted genome wide approach, although it probably excludes some genes of importance. This is the approach used for this work of thesis.

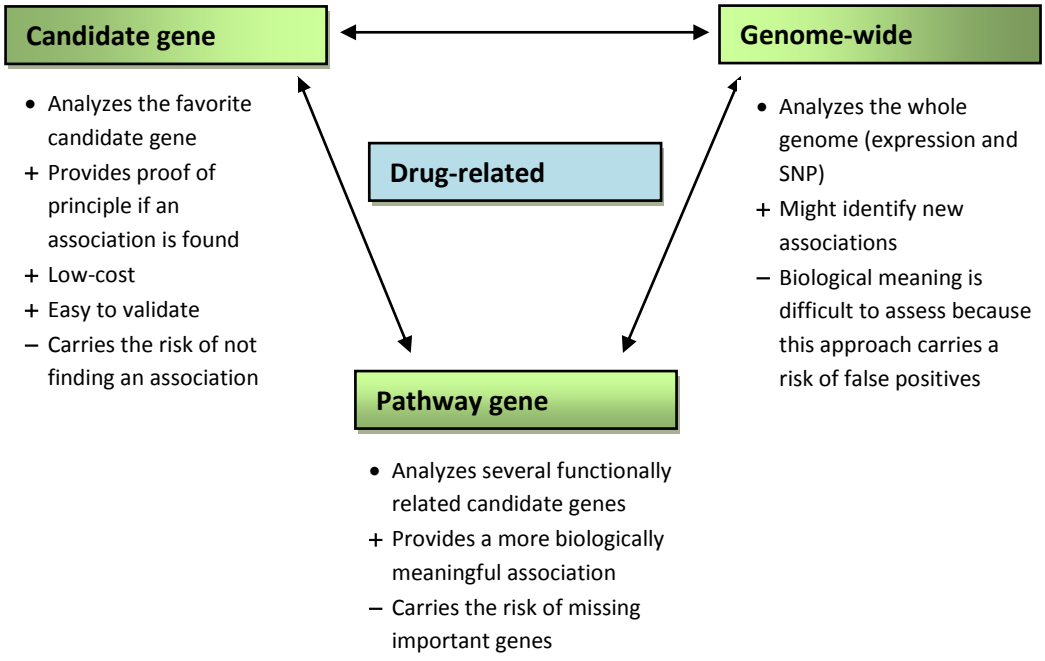


Figure 1. Pharmacogenetic approaches: differences among candidate-gene, pathway gene and genome wide approaches: main characteristics (•), advantages (+), and disadvantages (-).

The rationale for introducing genetic testing in clinical practice is that on one hand the most effective therapy permits to avoid both the waste of time connected to the choice of therapeutic unfit regimens, and the psychological negative relapses on patient due to prescribed therapy inappropriateness; on the other hand it is possible to assess a priori the risk of adverse events and to avoid to administer drugs or too high doses that could damage the patient.

However, pharmacogenetics has currently several limitations. The main one is that if a genetic marker is found in a patient population, it needs to be validated from a clinical point of view by a replication study in a different cohort of patients but with the same characteristics of the first one. Usually, there is a lack of clinical validation of obtained data due to the heterogeneity and sometimes to the conflicts that

emerges when the results reported in several studies are compared.

This discrepancy may be due to several factors (Efferth *et al.*, 2005; Ekhart *et al.*, 2009; Hoskins *et al.*, 2007):

- differences in study design (e.g. retrospective/prospective, low statistical power),
- survey unsuitable sample to evaluate the effect of low penetrance polymorphisms or to assess the importance of a haplotypic approach with the risk of generating false positives/negatives,
- differences in the clinical setting and treatment plan (e.g. dose and methods of administration of the drug, coadministration of other chemotherapeutic agents),
- heterogeneity in tumor pathology (e.g. stage of disease, tumor site, pre-treatment) and in clinical and demographic characteristics of patients (e.g. concomitant diseases, performance status, age, sex),
- inability to control environmental confusing factors (alcohol consumption, smoking, diet),
- differences in experimental techniques employed for the determination of polymorphisms and in parameters and methods of measurement of clinical outcome.

To identify efficient genetic markers of a specific chemotherapeutic treatment used for a particular pathology, it is important, therefore, to manage multicenter, methodologically well defined prospective pharmacogenetic studies using a population of patients adequately large and as uniform as possible from clinical, demographic and behavioral point of view. The integration of results from so structured pharmacogenetic studies would be an efficient strategy for reaching a tailored tumor therapy and for getting from each treatment the maximum effectiveness and the minimal toxicity, choosing the most suitable drug and the optimal dose for the individual.

1.2 BREAST CANCER

Breast cancer (BC) is the female most frequently diagnosed malignancy and the primary cause of cancer-related death among females worldwide (cancerresearchuk).

In 2008, 1.380.000 new cases and 458.000 BC deaths were world-wide registered and 332.000 new cases/89.000 deaths were reported in the European Union (Cardoso *et al.*, 2012).

Only in Italy, in 2011, 45.000 new breast cancer cases were diagnosed and 12.000 death have occurred, leading this disease becoming the first tumor-specific cause of death in the same year (Documento AIOM-AIRTUM, I numeri del cancro in Italia).

BC is a heterogeneous disease with several classifications. BC classification has the propose to select the best treatment accordingly to the tumor characteristics.

A full classification includes histopathological type, grade, stage (TNM), receptor status, and the presence or absence of genes as determined by DNA testing.

Histopathology. Histopathologic classification is based on characteristics seen when a light microscopy is applied on a biopsy specimen. The World Health Organization (WHO) has classified breast tumors according to histopathological features into several categories (Lakhani *et al.*, 2012), but the three most common types are:

- Invasive ductal carcinoma - 55% of BCs
- Ductal carcinoma in situ - 13% of BCs
- Invasive lobular carcinoma - 5% of BCs.

They collectively represent approximately three-quarters of BCs (Eheman *et al.*, 2009).

Grade. The tumor grading is based on the appearance of cancer cells and consists on determining how abnormal they and the tumor tissue look compared to normal cells and tissue.

BC can be classified as “well differentiated” (low-grade, grade 1), “moderately differentiated” (intermediate-grade, grade 2), and “poorly differentiated” (high-grade, grade 3 or 4), reflecting progressively less normal appearing cells that have a worsening prognosis. Indeed, the grade of a tumor is an indicator of how quickly it is likely to grow and spread in distant sites: the closer the appearance of cancer cells to normal cells, the slower their growth and the better the prognosis.

The Nottingham (also called Elston-Ellis) modification (Elston *et al.*, 2002) of the Scarff-Bloom-Richardson grading system (Bloom *et al.*, 1957; Genestie *et al.*, 1998) is recommended by the National Comprehensive Cancer Network (NCCN) guidelines Breast Cancer (Version 2.2011) which grades breast carcinomas by adding up scores for tubule formation, nuclear pleomorphism, and mitotic count, each of which is given 1 to 3 points. The scores for each of these three criteria are then added together to give an overall final score and corresponding grade as follows:

- 3-5 **Grade 1** tumor (**well-differentiated**), best prognosis.

- 6-7 **Grade 2** tumor (**moderately differentiated**), medium prognosis.
- 8-9 **Grade 3** tumor (**poorly differentiated**), worst prognosis.

Stage. The underlying purpose of staging is to describe the extent or severity of an individual's cancer, and to bring together cancers that have similar prognosis and treatment (https://cancerstaging.org/).

Staging information that is obtained prior to surgery, for example by mammography, x-rays and CT scans, is called *clinical staging* and staging by surgery is known as *pathological staging*.

Pathological staging is more accurate than clinical staging, but clinical staging is the first and sometimes the only staging type, for example in case of clinical stage IV disease, in which extensive surgery disease may not be helpful, and (appropriately) incomplete pathological staging information will be obtained.

The American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC) recommend TNM staging.

The TNM Classification of Malignant Tumors (TNM) is a cancer staging notation system that gives codes to describe the stage of a cancer, when this originates with a solid tumor.

- T describes the size of the original (primary) tumor and whether it has invaded nearby tissue,
- N describes nearby (regional) lymph nodes that are involved,
- M describes distant metastasis (spread of cancer from one part of the body to another).

TNM system is a two step procedure since first classifies cancer by T, N, M factors and then groups them into overall stages (I, II, III, IV).

Mandatory parameters for such a classification are listed below:

- **T**: size or direct extent of the primary tumor
    - Tx: tumor cannot be evaluated
    - Tis: carcinoma in situ
    - T0: no signs of tumor
    - T1, T2, T3, T4: size and/or extension of the primary tumor
- **N**: degree of spread to regional lymph nodes
    - Nx: lymph nodes cannot be evaluated
    - N0: tumor cells absent from regional lymph nodes
    - N1: regional lymph node metastasis present; (at some sites: tumor spread to closest or small number of regional lymph nodes)
    - N2: tumor spread to an extent between N1 and N3 (N2 is not used at all sites)
    - N3: tumor spread to more distant or numerous regional lymph nodes (N3 is not used at all sites)
- **M**: presence of distant metastasis
    - M0: no distant metastasis
    - M1: metastasis to distant organs (beyond regional lymph nodes).

Overall stages, which consider the TNM categorization, are reported in Table 2.

Stage	T	N	M
Stage 0	Tis	N0	M0
Stage IA	T1*	N0	M0
Stage IB	T0	N1mi	M0
	T1*	N1mi	M0
Stage IIA	T0	N1**	M0
	T1*	N1**	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1*	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

Table 2. Overall stages according to TNM system.

Receptor status. Today, an important determinant of treatment effects is the molecular characterization of BC. BC cells may or may not express three main receptors: estrogen receptors (ER), progesterone receptors (PgR) and human epidermal growth factor receptor 2 (HER2). Receptor status is a critical assessment for all breast cancers as it determines the suitability of using targeted treatments such as tamoxifen or other hormone therapies (e.g. aromatase inhibitors) and trastuzumab or lapatinib.

- **ER.** Cellular signaling of estrogens is mediated through two ERs, ERα and ERβ both belonging to the nuclear receptor (NR) family of transcription factors. Estrogen receptors are over-expressed in around 80% of BC cases. Estrogen receptor positive (ER+) cancer cells depend on estrogen for their growth, so they can be treated with drugs to reduce either the effect of estrogen (e.g. tamoxifen) or the actual level of estrogen (e.g. aromatase inhibitors), and generally have a better prognosis.
- **PgR.** Progesterone receptors are ligand-activated transcription factors that act in concert with intracellular signaling pathways as "sensors" of multiple growth factor inputs to hormonally regulated tissues, such as the breast. PgRs are useful prognostic indicators of BCs likely to respond to anti-estrogen receptor therapies (Lange *et al.*, 2008).
- **HER2**. HER2 is part of the epidermal growth factor (EGF) family and is overexpressed in around 20-30% of BC tumors. Prior to modern treatments, it was associated with a more aggressive disease, higher recurrence rate, and increased mortality, so a worse prognosis (Sotiriou *et al.*, 2009). However, HER2+ cancer cells respond to drugs such as the monoclonal antibody,

trastuzumab, (in combination with conventional chemotherapy) or the tyrosine kinase inhibitor lapatinib and this has significantly improved the prognosis (Romond *et al.*, 2005).

By a molecular point of view, BC consists of at least three different diseases: hormone-sensitive BC, the human epidermal growth factor receptor positive (HER2+) subtype, and triple-negative disease. Each molecular subtype has distinct biological features and a distinct clinical course: hormone receptor–positive (HR+) disease is generally characterized by a more indolent course, with a long disease-free interval (DFI) and a tendency to relapse in the bone or soft tissues; amplification of the HER2 gene confers a more aggressive clinical behavior to the HR+ subgroup, with a higher propensity for visceral relapses. Both triple-negative BC and hormone receptor–negative HER2+ BC are aggressive subtypes, with early visceral or central nervous system (CNS) metastases. Each molecular subtype requires distinct therapeutic approaches: in HR+ tumors, endocrine treatment is the cornerstone of therapy; in HER-2+ tumors, the incorporation of anti–HER2 agents has substantially reversed the negative prognostic impact of HER2 overexpression/amplification. Lastly, chemotherapy is the only available option so far for the triple-negative subtype, which is characterized by the absence of hormone receptors and HER2 negativity.

1.2.1 Molecular subtypes

Receptor status was traditionally evaluating by reviewing each individual receptor (ER, PgR, HER2) in turn, but nowadays approaches consider them together, along with the tumor grade, to categorize BC into conceptual molecular subtypes (Prat *et al.*, 2011) with different prognoses which may have different responses to specific therapies (Geyer *et al.*, 2009). Proposed molecular subtypes include:

- Basal-like: ER-, PgR- and HER2- or triple negative breast cancer (TNBC) (Perou *et al.*, 2011). Most triple negative tumors are basal-like and most basal-like tumors are triple negative but their overlapping is not complete. These tumors tend to occur more often in younger women. They are aggressive and have a poorer prognosis (at five years after diagnosis) compared to the ER+ subtypes (luminal A and luminal B tumors) (Carey *et al.*, 2006).
- Luminal A: ER+ and low grade: luminal A tumors tend to have the best prognosis, with fairly high survival rates and fairly low recurrence rates (Carey *et al.*, 2006; Metzger-Filho *et al.*, 2013). Because luminal A tumors tend to be ER+, treatment for these tumors often includes hormone therapy.
- Luminal B: ER+ but often high grade: Women with luminal B tumors are often diagnosed at a younger age than those with luminal A tumors (Metzger-Filho *et al.*, 2013; Lund *et al.*, 2010) and, compared to luminal A tumors, they tend to have poorer tumor grade, larger tumor size, lymph node-positive, p53 gene mutations (about 30%), all factors that lead to a poorer prognosis Carey *et al.*, 2006; Metzger-Filho *et al.*, 2013; Lund *et al.*, 2010). In some studies,

women with luminal B tumors have fairly high survival rates, although not as high as those with luminal A tumors (Carey *et al.*, 2006; Metzger-Filho *et al.*, 2013).

- HER2 type: ER-, PgR-, lymphnode-positive, poor tumor grade. About the 30% of HER2 type cancers do not exhibit HER2+ cells since HER2 type does not mean HER2+ cells. HER2 type tumors have a fairly poor prognosis and are prone to early and frequent recurrence and metastases. Women with HER2 type tumors appear to be diagnosed at a younger age than those with luminal A and luminal B tumors (Carey *et al.*, 2006; Voduc *et al.*, 2010).

Other less common molecular subtypes have also been described including:

- normal breast-like These tumors are usually small and tend to have a good prognosis (Carey *et al.*, 2006);
- apocrine molecular type or Luminal ER-/AR+: recently identified androgen responsive subtype which may respond to antihormonal treatment with bicalutamide (Lehmann *et al.*, 2011);
- claudin-low type: a very recently described class; often triple-negative, but different since there is a low expression of cell-cell junction proteins such as E-cadherin (Perou *et al.*, 2011), and frequently lymphocytes infiltrations are found (Prat *et al.*, 2011; Harrell *et al.*, 2012).

BCs that do not fall into any of these subtypes are often listed as unclassified.

1.3 METASTATIC BREAST CANCER

Despite the gains in early detection, approximately 5% to 10% of BCs are metastatic at diagnosis; of these, approximately one-fifth will survive 5 years (Cardoso *et al.*, 2012).

In addition, in spite of advances in treatment strategies about 25%-40% of patients with early-stage, non-metastatic BC at diagnosis will develop distant metastatic disease, that is largely incurable, so therapeutic decisions should be realistic and patient-specific (Early Breast Cancer Trialists' Collaborative Group, Cardoso *et al.*, 2012; Guarnieri *et al.*, 2009; Pagani *et al.*, 2010).

Metastatic breast cancer (MBC) is a heterogeneous disease that has a variety of different clinical scenarios, ranging from solitary metastatic lesions to diffuse and multiple organ involvement. Although MBC is unlikely to be cured, there have been meaningful improvements in survival due to the availability of more effective systemic therapies (Giordano *et al.*, 2004; Chia *et al.*, 2007; Gennari *et al.*, 2005; Dawood et al., 2008, Swain *et al.*, 2013).

In 1996, median overall survival (OS) approached two years, with a range from a few months to many years (Greenberg *et al.*, 1996). Now, the median survival for MBC is approximately 18 to 30 months, this means that survival of patients with MBC is slowly but steadily improving and the risk of death is decreasing by 1%–2% each year (Pagani *et al.*, 2010).

The greatest improvement is most probably related to the development and widespread availability of modern systemic therapies. In addition, modern diagnostic tools allow the detection of early metastatic disease, which may be more responsive to treatment than late metastatic disease.

Treatment goals vary from symptom control to lengthening survival, mainly on the basis of patient age and

performance status, tumor biology, site and extent of disease, and prior therapies (Guarnieri *et al.*, 2009).

The selection of a therapeutic strategy depends upon both tumor biology and clinical factors, with the goal being a tailored approach. Although a subset of patients with oligometastatic disease may benefit from an intensified locoregional treatment, most patients with MBC receive systemic medical therapy consisting of chemotherapy, endocrine therapy, and/or biologic therapies, and supportive care measures (Pagani *et al.*, 2010).

1.3.1 Aromatase inhibitors therapy for hormone sensitive MBC patients

Estrogens are the female sexual hormones that promote the formation of secondary sexual characteristics, such as breast, the enlargement of the basin and are involved in the proliferation of the endometrium and various phenomena of the menstrual cycle.

The main actors in the estrogens scenario are three: estradiol (E_2), estriol (E_3) and estrone (E_1). From

menarche to menopause the most important estrogen is 17 β-estradiol (E_2); after menopause, instead, the estrone is mainly produced estrogen and it has a lower activity than E_2.

Estrogens are produced by the ovarian follicles and the placenta. Their synthesis starts in the theca interna cells of the ovaries: here, cholesterol is converted into androstenedione, a precursor with low estrogenic activity. The androstenedione therefore, crosses the basal membrane and reaches the granulosa cells, where it is converted into estrone or estradiol.

Some estrogens are produced in other organs and tissues in small amounts. This peripheral production represents the only source of estrogen for men and for women in menopause. In fact, despite the menopause results in the cessation of ovarian estrogen synthesis, low estrogen concentrations continue to be present in women plasma. Initially it was believed that these hormones were produced by the adrenal glands, but later it became clear they only produce circulating androgens, which are then converted into estrogen in different tissues, such as bone, liver, muscle, adipose tissue and breast tissue both benign and malignant.

Responsible for the estrogenic conversion of androgens is the aromatase, an enzyme belonging to the cytochrome P450 superfamily and encoded by the gene *CYP19A1*.

Ovaries, placenta, brain tissue, bone, fat, and breast (Brueggemeier *et al.*, 2006) expresses aromatase, which catalyzes the last "step" of estrogen synthesis by the conversion of androstenedione and testosterone into estrone and estradiol, respectively (Stanczyk *et al.*, 2003). The catalytic action of aromatase takes place through the oxidation and the elimination of a methyl group from the A ring of androgens that is converted into an aromatic ring. The enzyme is named by the reaction of aromatization, which characterizes its catalytic activity (Fig. 2).

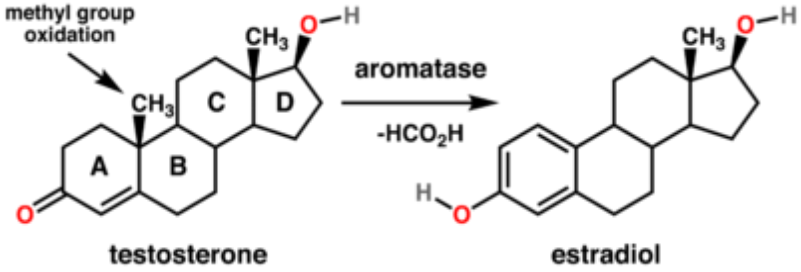


Figure 2. Testosterone A ring aromatization reaction catalyzed by aromatase enzyme. The final product is an aromatic derivative.

The major pathway of estrogen local synthesis is the aromatization of androstenedione into estrone. The estrone by itself does not stimulate the activation of the estrogen receptor but it is easily converted into estradiol by multiple dehydrogenases. During the fertile age of the woman, the concentrations of estrone are lower than those of estradiol, the latter predominating also regarding the action power. The estrone, which becomes predominant only after menopause due to stoppage of the ovarian estradiol production, can be converted into estrone sulfate (E_1S), a derivative with a prolonged half-life, which acts as a reserve of estrone and estradiol, produced at cellular level by the steroid-sulfatase enzyme. To

be biologically active, in fact, E_1S must be deconjugated from sulfate and subsequently be converted into estradiol (Lønning *et al*., 2013).

Thanks to aromatase, therefore, estrogens are produced directly in the benign and malignant breast tissue, stimulating the growth of ER+ cells (Brueggemeier *et al.*, 2005). This biochemical process is the biological rationale for the development of aromatase inhibitors, a class of compounds widely used in the treatment of postmenopausal BC (Buzdar *et al.*, 2002).

Aromatase inhibitors are generally divided into two major classes: non-steroidal aromatase inhibitors (NSAI) and steroidal inhibitors (Fig. 3).

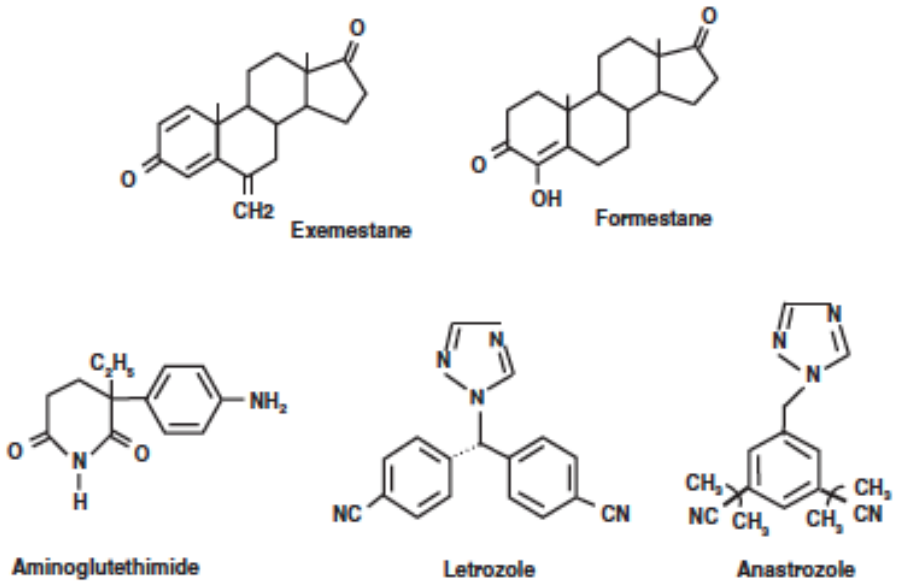


Figure 3. Steroidal and non-steroidal aromatase inhibitors.

Non-steroidal inhibitors belong to two different chemical classes: the aminoglutethimide analogues (aminoglutethimide and rogletimide) and the triazole derivatives (anastrozole, letrozole and vorozole). The steroidal compounds include formestane (second generation) and exemestane (third generation), both are androstenedione derivatives, the main aromatase substrate.

These two classes differ in their biochemical action exerted on aromatase. In fact, while the non-steroidal inhibitors form with the enzyme a reversible binding that involves the iron/heme group of the cytochrome, steroidal inhibitors compete with the endogenous substrates (androstenedione and testosterone) to the active site of the enzyme in an irreversible manner: for this reason they are also referred to as "suicide inhibitors" or aromatase inactivators (Brueggemeier *et al*., 2005).

It has been hypothesized that the non complete cross-resistance observed between steroidal and non-steroidal compounds is attributable to their different action on the enzyme, but has also been supposed that it can involve additional endocrine effects of steroid compounds (Lønning *et al*., 2004).

1.3.2 Exemestane

Exemestane is a third generation steroidal aromatase inhibitor that, because of its irreversible binding, determines the inactivation of the enzyme. Exemestane is used in adjuvant setting for treating postmenopausal women with ER+ invasive BC at early stage. The drug is also indicated in the treatment of advanced stage BC, when the disease has progressed following an anti-estrogen therapy.

Several studies have evaluated two different therapeutic approaches, the first involving the use of exemestane monotherapy, the other a sequential administration of exemestane after 2-3 years of tamoxifen treatment. The rationale for the sequential approach was based on the results obtained from the treatment of MBC patients that highlighted the lack of cross-resistance between tamoxifen and aromatase inhibitors (Lønning *et al*., 2013).

Coombes and colleagues (Coombes *et al*., 2004) have conducted a randomized trial on 4724 patients in order to evaluate the efficacy of switching to exemestane after 2-3 years of tamoxifen for a total of 5 yearlong antiestrogen therapy versus 5 years of tamoxifen alone. The results of this huge study suggest that early improvements in disease-free survival noted in patients who switch to exemestane after 2-3 years on tamoxifen persist after treatment, and translate into a modest improvement in OS (Coombes *et al*., 2007).

Other clinical studies investigating the use of exemestane in BC adjuvant treatment have shown greater efficacy of exemestane in reducing the contralateral cancer (56% reduction with exemestane versus tamoxifen). In the study of letrozole versus placebo reduction of contralateral cancer was 46% (Goss *et al.*, 2003).

The scenario of the use of aromatase inhibitors in MBC setting has changed in recent years, since nowadays many patients with a ER+ progressing MBC have already received at least an aromatase inhibitor, generally a non-steroidal one, in the adjuvant setting.

Hence, guidelines recommends that for patients relapsing after more than 12 months by the conclusion of adjuvant therapy, re-use of an aromatase inhibitor can be a reasonable choice. In contrast, patients who relapse during treatment or shortly after its conclusion need an alternative treatment option (Jurado *et al.*, 2011).

Exemestane, being a steroidal inhibitor, has shown to be effective in metastatic setting even after failure of a prior treatment with non-steroidal aromatase inhibitors. In fact, a phase II clinical trial in which exemestane antitumor activity and its toxicity were evaluated in patients with MBC progressed after a non-steroidal aromatase inhibitors treatment, has shown that exemestane was effective in terms of objective response, thus demonstrating to be a valid therapeutic strategy for patients who become resistant to non-steroidal aromatase inhibitors (Lønning *et al*., 2000).

In the Evaluation of Faslodex (fulvestrant) versus Exemestane Clinical Trial (EFECT), a randomized study by Chia and collaborators (Chia *et al.*, 2008) a similar efficacy between fulvestrant and exemestane has been demonstrated in a cohort of 693 patients with locally advanced BC in which anastrozole or letrozole had previously failed.

The SoFEA study shown similar results on 723 patients, in which the combinations fulvestrant + anastrozole or fulvestrant + placebo was not found to be better than either fulvestrant or exemestane alone in terms of Progression Free Survival (PFS) in postmenopausal women with HR+ advanced BC, after the loss of response to NSAIs (Johnston *et al.*, 2013).

Exemestane demonstrated also a greater efficacy than megestrol acetate (PgR agonist) in patients with MBC refractory to tamoxifen, both in terms of median time to progression (TTP) and median survival.

Data from phase III studies, comparing exemestane with tamoxifen in the first-line therapy of MBC patients (Coombes *et al.*, 2004, Paridaens *et al.*, 2008) confirm the phase II study demonstrating the superiority of exemestane compared tamoxifen (Paridaens *et al.*, 2003).

1.3.3 Exemestane toxicity

Osteoporosis is a major problem that affects the elderly female population in many countries, as osteoporotic fractures are associated with significant morbidity and excessive mortality (Lønning *et al*., 2013). Estrogens, indeed, play a key role in many physiological processes other than the reproduction. Estrogen deprivation may result in a wide range of effects as demonstrated by aromatase gene knock-out mice which, not being able to produce estrogen, reveal multiple metabolic defects and especially osteopenia. Today it is widely accepted that inhibition of aromatase in early BC produces a moderate increase in bone loss and disturbances in lipid metabolism that may increase the risk of cardiovascular disease. It is, however, necessary a distinction between the effects caused by NSAI and the steroidal ones: many studies, in fact, have confirmed that non-steroidal inhibitors increases the rate of bone turnover, accelerate bone loss and increase the incidence of fractures (Coleman *et al.*, 2007). Instead, in preclinical models in rats, exemestane and its metabolite 17-OH exemestane have shown, in addition to a significant reduction in total and LDL cholesterol levels, a protective action in bone loss after ovariectomy (Goss *et al.*, 2004). These results, however, are only partially confirmed in humans. In fact, with regard to lipid metabolism, Lønning and collaborators show only a 6% to 9% reduction in HDL cholesterol levels (but not LDL) without any other effect on serum lipids, coagulation factors or homocysteine levels (Lønning *et al.*, 2005).

Regarding the comparison between exemestane and tamoxifen toxicity, studies are conflicting. In the Integroup Exemestane Study (IES) conducted on 4724 patients randomized to receive tamoxifen or examestane treatment, toxic effects of any grade were reported in 92.5% of patients treated with

exemestane and in 92.6% of those treated with tamoxifen. Grade 3 or 4 adverse events were experienced by the 18.4% of patients in the exemestane arm compared to 17.6% in the tamoxifen arm. In this study, exemestane was associated with a higher incidence of arthralgia, musculoskeletal pain, carpal tunnel syndrome, joint stiffness, paresthesia, fractures, arthritis, osteoporosis, diarrhea, headache and insomnia while gynecological symptoms, vaginal bleeding, endometrial hyperplasia, fibroids, uterine polyps and cramps were more common among patients treated with tamoxifen. Thromboembolic events were more frequently associated with tamoxifen (1.2% vs 2.3%) and cardiovascular events does not seem to differ between the two groups, although a non significant trend was highlighted for myocardial infarctions with 1.3% of cases in the exemestane group and 0.8% of patients in the tamoxifen group. Hot flashes were reported as toxic effect in 41.3% of patients treated with exemestane versus 38.6% in patients in the tamoxifen arm (Coombes *et al.*, 2007). Paridaens and colleagues in the phase III study comparing exemestane with tamoxifen as first-line hormonal treatment of MBC in 317 postmenopausal women reported slightly different results. Exemestane has proven to be more effective than tamoxifen (PFS 9.9 months vs. 5.8) and to have good tolerability. Hot flashes were experienced only by the 6.5% of patients treated with exemestane versus the 12.2% of patients treated with tamoxifen but patients in the exemestane arm had more grade 1/2 arthralgia/myalgia, more cardiac dysrhythmia, cardiac dysfunction, and more grade 1 diarrhea (Paridaens *et al.*, 2008).

1.3.4 Pathway

The estrogenic synthesis starts from androstenedione or testosterone which are produced by several enzymatic steps, one of which is catalyzed by cytochrome P450 17A1, or steroid 17-alpha-monooxygenase, or 17α-hydroxylase/17,20 lyase is encoded by the CYP17A1 gene (Fig. 4).

CYP17A1 is a key enzyme in the steroidogenic pathway that produces progestins, mineralocorticoids, glucocorticoids and androgens such as the androstenedione. Androstenedione can be converted, by the action of the 17β-Hydroxysteroid dehydrogenases, into testosterone.

Both androstenedione and testosterone are substrate for the aromatase enzyme (encoded by CYP19A1 gene), which, as previously described, converts them into estrone and 17 β estradiol by the aromatization of the steroidal nucleus.

17 β estradiol activates the estrogen receptors ER1 and ER2, encoded by *ESR1* and *ESR2*. The interaction of 17 β estradiol and its receptors leads to the receptors dimerization and their translocation into the cell nucleus where, by recognizing the estrogen responsive elements (EREs) on the DNA sequence, they trigger several growth signals.

Retinoblastoma-interacting zinc-finger protein 1 (RIZ1) an estrogen receptor alpha (ESR1) specific co-activator, strongly enhance ESR1 function.

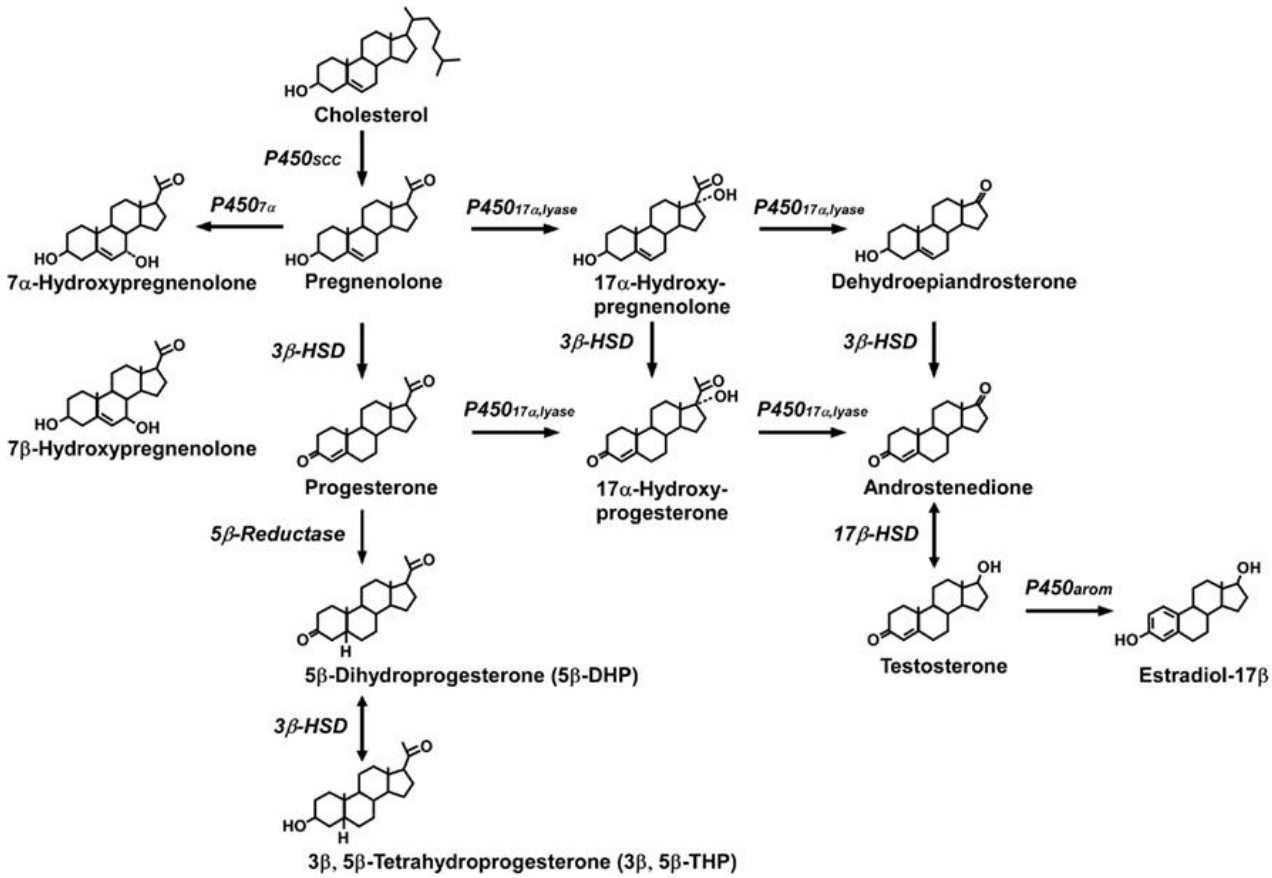


Figure 4. Estrogen biosynthesis.

Estrogens are metabolically converted into hormonally inactive (or less active) water-soluble metabolites firstly by oxidative reactions (largely hydroxylations (Martucci *et al.*, 1993)) and later by conjugative metabolism (glucuronidation, sulfonation, and *O*-metilation).

Members of the cytochrome P450 family are the major enzymes catalyzing the oxidative metabolism of estrogens to multiple hydroxilated metabolites. (Zhu and Conney., 1998)

CYP1B1 is a heme-thiolate monooxygenase, encoded by the *CYP1B1* gene, which mainly catalyzes the 2- and 4-hydroxylation of 17 βestradiol (2-OH E_2, 4-OH E_2)(Gajjar *et al.*, 2012).

Hydroxylated estrogens undergoes a phase II conjugation metabolism exerted by the uridine diphospho glucuronosyltransferase 1A1 (UGT1A1) and the Catechol-*O*-methyl transferase (COMT), encoded by *UGT1A1* and *COMT* genes, respectively.

Conjugates estrogens does not bind the ERs, so they are mostly considered inactive metabolites (Zhu and Conney 1998). Estrogen glucuronidation is catalyzed by several members of the UDP glucuronosyltransferase (UGT) superfamily enzymes (Albert *et al.*, 1999; Meech and Mackenzie, 1997). UGTs catalyze the conjugation of UDP-glucuronic acid to a variety of endogenous and exogenous aglycones, including steroid hormones. Whereas estrogens are sulfated predominantly at the 3 position, glucuronidation can occur at either the 3 or 17 hydroxyl group of steroidal hormones, with the 17

position being the apparent preferred site of glucuronidation for 17-estradiol. Glucuronidation of estrogens renders

those molecules less lipophilic and more readily excreted in both urine and bile (Raftogianis *et al.*, 2000). The O-methylation of cathecol estrogens (2-OH E_2 and 4 OH E_2) is catalyzed by the COMT, an enzyme that catalyzes the transfer of a methyl group from the methyl donor S-adenosyl methionine (SAM) to one hydroxyl moiety of the catechol ring of a substrate (Dawling *et al.*, 2001). The resulting monomethylated estrogens (2-MeO E_2; 2-OH-3-MeO E_2; 4-MeO E_2; 4-OH-3-MeO E_2)have little or no estrogen receptor binding affinity but have long half-lives (Zhu and Conney, 1998).

Exemestane inactivate the aromatase gene and its action is translated into an estrogen synthesis inhibition.

Exemestane is extensively metabolized by the CYP3A4 enzyme at hepatic level.

The initial steps in the metabolism of exemestane are oxidation of the methylene group in position 6 and reduction of the 17-keto group with subsequent formation of many secondary metabolites. The metabolites are inactive or inhibit aromatase with decreased potency compared with the parent drug. (Aromasin labeling by Pfizer: http://labeling.pfizer.com/ShowLabeling.aspx?id=523).

1.3.5 Pharmacogenetics of exemestane

The studies so far conducted seem to indicate a better toxicity profile and clinical efficacy of exemestane than tamoxifen both in adjuvant and in advanced setting (Coombes *et al.*, 2007; Paridaens *et al.*, 2008). Exemestane has also a more selective specificity of action toward BC cells than other aromatase inhibitors (Brueggemeier *et al.*, 2005). However, in some patients, exemestane therapy may be not effective (primary or *de novo* resistance) and its toxic effects may sometimes require the treatment suspension (Coombes *et al.*, 2004). Moreover, many patients experience disease progression regardless of an early clinical benefit (CB) (acquired resistance).

Patient-related clinical-pathological factors may influence aromatase inhibitors activity. Several biological mechanisms have been proposed to explain resistance to endocrine therapy ranging from epigenetic mechanisms to alterations in cellular signaling pathways (deregulations of the ER pathway, alterations in cell cycle and apoptotic machinery, activation of alternative pathways conferring cell survival and proliferation). In particular, genetic variations of the aromatase gene, or genes involved in AIs and estrogens activity and metabolism, have been suggested too (Lazarus *et al.*, 2010; Sun, 2010).

The wide range of responses to exemestane therapy observed within a population of patients may be due to an impaired activity of the aromatase target enzyme or to an inter-patient difference in the metabolic elimination of exemestane, leading to variations in the pharmacokinetic profile of the drug. Moreover, inter-individual differences in the biosynthesis and catabolic pathways of estrogens, as well

as changes in the expression of their cellular receptors and transporters, could directly affect the pharmacodynamics (toxicity and response) and the ability to have an appropriate estrogen suppression mediated by exemestane.

Genetic polymorphisms have been described for all the biochemical and molecular events that can potentially interfere with the pharmacokinetic and pharmacodynamic effects (toxicity and response) of exemestane therapy (Fig. 5). Therefore, the study of these polymorphisms may be useful to better understanding of the response to exemestane observed in certain individuals. Genetic analysis can be employed for exemestane therapy personalization, that is, in the definition of which individuals are more susceptible to have a better response or a specific toxic effects and, accordingly, in the planning of dosage variations in case of ineffectiveness due to differences in drug metabolism or target.

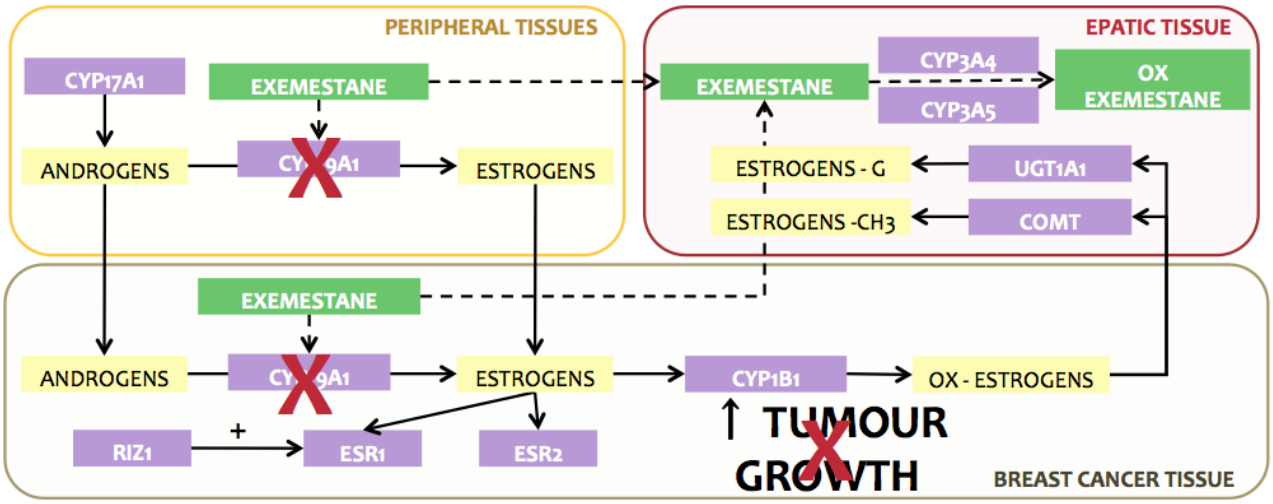


Figure 5. Schematic representation of the estrogen biosynthesis, catabolism and mechanism of action. Exemestane mechanism of action and metabolism.

Several polymorphisms have been described as potentially involved in the efficacy of exemestane (Table 3). In particular, polymorphisms in the aromatase gene could be linked to variations in the systemic estrogen availability and a different ability to interact with the drug. Colomer and collaborators (Colomer *et al.*, 2008) reported that patients treated with letrozole and carriers of the homozygous variant genotype for CYP19A1_Ex11+410A/C (rs4646) SNP in the 3'UTR, had a better TTP compared to patients with wild-type genotype. This, like other polymorphisms, such as the variable number of repeats CYP19A1_(TTTA)n (rs60271534) in intron 4 of the gene, the CYP19A1_1558T/C (rs10046) SNP in exon 10 of the 3'UTR, and the CYP19A1_47T/C (rs700519) SNP in exon 8 (Arg264Cys) have been linked to a decrease in the expression or activity of the enzyme and may, therefore, influence the effectiveness of the drug.

These polymorphisms have been also associated with changes in bone mineral density in postmenopausal women, emphasizing their phenotypic impact and a possible involvement in individual

predisposition to develop side effects to the drug. (Tofteng *et al.*, 2004; Mendoza *et al.*, 2008; Yilmaz *et al.*, 2011). The nonsynonimous Arg264Cys CYP19A1_47T/C (rs700519) SNP has been also associated with BC survival in a cohort of 1136 BC Chinese patients (Long *et al.*, 2006).

PATHWAY	GENE	SNP	RS ID	EFFECT	ASSOCIATIONS
E synthesis, Exe activity	*CYP19A1*	Ex11+410A/C	rs4646	↓ activity/expression	↑ TTP[1]
E synthesis, Exe activity	*CYP19A1*	(TTTA)n	rs60271534	↓ activity/expression	Bone mineral density[2]
E synthesis, Exe activity	*CYP19A1*	1558T/C	rs10046	↓ activity/expression	Bone mineral density[2]
E synthesis, Exe activity	*CYP19A1*	47T/C	rs700519	↓ activity/expression	BC survival[3]
E synthesis	*CYP17A1*	27A/G	rs743572	unknown[a]	Osteoporosis[4]
E activity	*ESR1*	256A/G (Xbal)	rs9340799	↓ ESR1 mRNA	Osteoarthritis[6]
E activity	*ESR1*	497T/C (Pvull)	rs2234693	↓ activity/expression	Osteoporosis[5]
E activity	*ESR2*	1082A/G	rs1256049	unknown	Prostate cancer risk[7]
E activity	*ESR2*	1730A/G	rs4986938	↓ ESR2 mRNA[8]	BC Survival[9]
E activity	*RIZ1*	delP704	rs2308040	unknown	Bone mineral density[10]
E phase I metabolism	*CYP1B1*	*3_4326G/C	rs1056836	↑ activity [11]	Endometrial cancer risk[12]
E phase II metabolism	*UGT1A1*	*28_TA(6/7)	rs8175347	↓ transcription rate[13]	BC risk[14]
E phase II metabolism	*COMT*	12A/G	rs4680	↓ stability/activity[15]	BC risk[16]
Exe metabolism	*CYP3A4*	*1B	rs2740574	↑ transcription rate[21]	Steroid metabolism[22]
Exe metabolism	*CYP3A5*	*3	rs776746	Non fuctional protein[22]	Steroid metabolism[22]

Table 3. List of the investigated SNPs according to the biological mechanism they are involved in (E = estrogens; Exe = exemestane) along with their functional effects and already known associations.
[1.]Colomer *et al.*, 2008; [2.]Tofteng *et al.*, 2004; [3.]Long *et al.*, 2006; [4.]Somner *et al.*, 2004; [5.]Ioannidis *et al.*, 2004; [6.]Yin *et al.*, 2014; [7.]Fu *et al.*, 2014; [8.]Putnik *et al.*, 2009; [9.]Borgquist *et al.*, 2013; [10.]Grundberg *et al.*, 2004; [11.]Shimada *et al.*, 1999; [12.]Teng *et al.*, 2013; [13.]Bosma *et al.*, 2003; [14.]Adegoke *et al.*, 2004; [15.]Lachman *et al.*, 1996; [16.]Wan *et al.*, 2014.
A Olson et al., 2007;

The CYP17A1 is another enzyme involved in the synthesis of steroid hormones after menopause. In particular, this enzyme is responsible for the synthesis of androstenedione, an estrogenic precursor. The polymorphism CYP17A1_27A/G (rs743572) in exon 1 of the 5'UTR of the gene has been linked to osteoporosis and, as polymorphisms of *CYP19A1*, may play a role in the efficacy or be associated with estrogen dependent side effects of exemestane (Somner *et al.*, 2004).

It was reported that changes in the structure of estrogen receptors could in turn influence the effectiveness of estrogens, and then the activity of exemestane. In particular, for the Estrogen Receptor 1 (ER1), encoded by the *ESR1* gene, were considered the following polymorphisms:

- ESR1_497T / C (rs2234693), intronic,
- ESR1_256A / G (rs9340799) in exon 1

For the Estrogen Receptor 2 (ER2), encoded by the *ESR2* gene the following polymorphisms were analyzed:

- ESR2_1082A / G (rs1256049) in exon 5,
- ESR2_1730A / G (rs4986938) in the 3'UTR of exon 8;

These polymorphisms were associated with variations in the expression levels of receptor and in the reduction of bone density in postmenopausal women (Scariano *et al.*, 2004; Ioannidis *et al.*, 2004), a phenomenon exacerbated by the use of aromatase inhibitors.

It was also described a polymorphism in the *RIZ1* gene encoding for a potent coactivator that greatly enhances the activity of ESR1. The polymorphism RIZ1_delP704 (rs2308040) leads to the proline deletion at position 704 with a detrimental effect on bone density (Grundberg *et al.*, 2004).

Estrogens exposure, antagonized by exemestane, is also influenced by the catabolic steps mediated by CYP1B1, involved in estradiol oxidation, and UGT1A1 and COMT, involved in conjugation and in detoxification of estradiol and his oxidized metabolites.

It has been shown that polymorphisms affecting the expression or activity of these proteins have a role in hormone-dependent tumors such as breast and endometrial cancers. This may be due to alterations in exposure of breast tissue to the trophic effects of estrogen or to the mutagenic effects of their oxidized metabolites (Zhu and Conney, 1998). Several polymorphisms have been described for this class of enzymes. In particular, CYP1B1*3_4326G/C (rs1056836), UGT1A1*28_TA(6/7) (rs8175347) and COMT_12A/G (rs4680) in exon 4 were previously associated with the development of breast and endometrial cancer (Guillemette *et al.*, 2000; Zheng *et al.*, 2000). Changes on estrogen availability in breast tissue could affect the therapeutic effect of exemestane and the patients probability to relapse. Furthermore, the analysis of these polymorphisms could define patients who, by having a lower efficiency in detoxification or an estrogen receptor increased sensitivity, may be less responsive to the drug and require an increase in the dosage.

Some specific side effects of aromatase inhibitors (cardiovascular alterations, lipid metabolism disorders or bone loss) may be a consequence of an altered metabolism of exemestane or biochemical-molecular specificity characterizing individuals (Tofteng *et al.*, 2004; Dick *et al.*, 2005).

Exemestane undergoes an oxidative metabolism mainly related to CYP3A4. A polymorphism that affects the promoter of the gene (CYP3A4*1B - rs2740574) was associated with changes in the expression of the protein (Dai *et al.*, 2001), which could lead to changes in the pharmacokinetics of exemestane.

It has been finally supposed the possible involvement of CYP3A5 in the metabolism of exemestane, as it catalyzes the oxidation of substrates very similar to those of CYP3A4. CYP3A5*3 intronic polymorphism (rs776746) is related to a change in the expression levels of the protein (Kuehl *et al.*, 2001).

Therefore, in this work of thesis, polymorphisms involved in the activity of aromatase, the mechanism of action and metabolism of estrogen and finally metabolism of exemestane were analyzed. A preventive screening, based on the individual pharmacogenetic profile, could be useful to identify patients with a better chance of responding to exemestane treatment or patients with a higher risk of developing side effects during the anti-aromatase therapy.

2. RATIONALE

In a sub-set of BC patients, exemestane as metastatic first-line treatment, appears to have a limited effectiveness. Exemestane inhibits the estrogens synthesis, so proteins cooperating to the estrogens availability of and their activities could have a role in the efficacy of exemestane therapy. Polymorphisms characterizing genes involved in estrogen availability, in the pharmacokinetics and pharmacodynamics of exemestane could alter the ability to produce the adequate estrogen suppression mediated by the drug.

In particular, polymorphisms on the *CYP19A1* gene, the drug's target, may influence the enzyme activity and thus the tumor response to the treatment. The homozygous variant genotype of the CYP19A1_Ex11_410A/C (rs4646) SNP in the 3'UTR of the aromatase gene, was related to a better disease-free survival (DFS) (Colomer *et al.*, 2008) in patients treated with letrozole and to a better OS in a population 272 Chinese patients treated with anastrozole (Liu *et al.*, 2013). Other polymorphisms in genes involved in the estrogens metabolism or activity may modulate the pharmacological efficacy of exemestane. These include polymorphisms on estrogens receptors (*ESR1*: ESR1_497T/C (rs2234693) and ESR1_256A/G (rs9340799); *ESR2:* ESR2_1082A/G (rs1256049) and ESR2_1730A/G (rs4986938); *RIZ1:* RIZ1_delP704 (rs2308040)), on the enzymes responsible for estrogens synthesis (CYP19A1_ (TTTA)n (rs60271534) in intron 4 of the gene, CYP19A1_1558T/C (rs10046) and CYP19A1_47T/C (rs700519) in exon 8; CYP17A1_27A/G (rs743572) in the 5'UTR of the gene) or their catabolism (CYP1B1*3_4326G/C (rs1056836), UGT1A1*28_TA(6/7) (rs8175347) and COMT_12A/G (rs4680) in exon 4). Finally, the enzymes involved in the oxidative metabolism of exemestane, mediated by CYP450 isoforms with polymorphic characteristics (CYP3A4*1B_-392A/G (rs2740574) and CYP3A5*3_6986A/G (rs776746)), could have an impact in determining the success of the exemestane-based therapy.

A preventive screening, based on the individual pharmacogenetic profile, might be useful to identify patients who are more likely to be responsive to exemestane. The safety profile of exemestane is quite good, however, in some patients the therapy discontinuation due to toxicity is necessary. A pharmacogenetic monitoring may help to determine which patients are more likely to develop estrogen-dependent side effects during the anti aromatase therapy.

3.AIMS

Since drug response and toxicity may be affected by patients' genetic background, Pharmacogenetics could represent an innovative tool to tailor cancer treatment in order to obtain the best drug effect and save patients from unnecessary toxicity.

In this PhD thesis, several germ-line polymorphisms in genes related to exemestane mechanism of action and metabolism or in estrogen biosynthesis, activity and catabolism have been investigated in a cohort of postmenopausal metastatic or locally advanced inoperable ER+ breast cancer (BC) patients treated with exemestane as first-line hormone therapy included in a prospective clinical study.

The main objectives of this thesis were the following:

1. Development of analytical methods for the characterization of genetic polymorphisms
2. Evaluate the effect on response rate (RR) and clinical benefit (CB) of the genetic polymorphism CYP19A1_Ex11+410A/C (rs4646) in the 3'UTR of the aromatase gene (exemestane target), which represent the primary genetic end point of the clinical study.
3. Evaluate the effect of genetic polymorphisms different from CYP19A1_Ex11+410A/C (rs4646) on the RR, CB and time to progression (TTP) to exemestane treatment:
    a. Other *CYP19A1* gene polymorphisms: CYP19A1_(TTTA)n (rs60271534) in intron 4, CYP19A1_1558T/C (rs10046) and CYP19A1_47T/C (rs700519) in exon 8.
    b. Polymorphisms involved in:
        i. Estrogen biosynthesis: CYP17A1_27A/G (rs743572) in the *CYP17A1* gene 5'UTR.
        ii. Estrogen catabolism and detoxification pathway: CYP1B1*3_4326G/C – Leu432Val (rs1056836) of *CYP1B1* gene; UGT1A1*28_TA(6/7) (rs8175347) of *UGT1A1* gene and COMT_12A/G – Val158Met (rs4680) in the exon 4 of *COMT* gene.
    c. Polymorphisms on estrogen receptors ERα and ERβ encoded by *ESR1* and *ESR2*, respectively:
        i. *ESR1*: ESR1_497T/C (rs2234693)and ESR1_256A/G (rs9340799);
        ii. *ESR2*: ESR2_1082A/G (rs1256049)and ESR2_1730A/G (rs4986938);
        iii. *RIZ1*, *ESR1* coactivator: RIZ1_delP704 (rs2308040), a proline deletion in position 704 of the protein.
    d. Polymorphisms of CYPs involved in the metabolism of exemestane: CYP3A4*1B_-392A/G (rs2740574) and CYP3A5*3_6986A/G (rs776746) of the *CYP3A4* and *3A5* genes, respectively.
4. Assess the effect of all the above-cited polymorphisms on safety profile (toxicity) of exemestane and overall survival (OS).

4. MATERIALS AND METHODS

4.1 PATIENTS ENROLMENT AND TREATMENT

This study, sponsored by the CRO-National Cancer Institute of Aviano, Italy, includes prospectively enrolled metastatic or locally-advanced BC patients which satisfied the following eligibility criteria:

1. Post menopausal metastatic or locally-advanced inoperable HR positive (ER and PgR) BC women at the time of study entry. ER or PgR were referred as positive in the presence of at least 10% nuclear staining in tumor cells;
2. Patients with recurrent BC after receiving adjuvant chemotherapy and/or hormonal therapy (tamoxifen, LHRH agonists, ovariectomy, non steroidal AIs);
3. Measurable lesion as defined by RECIST criteria version 1.0 (Therasse *et al.*, 2000): i.e. lesions that can be accurately measured in at least one dimension with the longest diameter ≥20 mm using conventional techniques or ≥10 mm using spiral CT scan. Signed informed consent and local IRB approval were requested;
4. No target lesions as defined by RECIST criteria were also considered;
5. Patients with advanced BC treated with first line hormone therapy whose disease had progressed following tamoxifen therapy;
6. Age ≥ 55 years;
7. Performance status (ECOG) 0 to 2;
8. Life expectancy >3 months;
9. Absolute neutrophil count (ANC) ≥1,500 μl, platelets ≥100,000 μl; hemoglobin ≥9.0 g/dl.

Exclusion criteria were:

1. Prior exemestane treatment;
2. Prior hormonal therapy for the metastatic setting;
3. Documented cerebral metastasis;
4. Serious active infectious disease;
5. Serious functional alteration of visceral and metabolic disease;
6. Radiotherapy or major surgery within 4 weeks from beginning of exemestane therapy;
7. Presence of previous or concomitant neoplasm with exclusion of in situ cervical cancer;
8. Patients who could not attend periodic clinical check-ups.

Dosage, schedule, and duration of treatment was based on exemestane current clinical setting. Patients were treated with exemestane, 25-mg/p.o./single daily dose. The planned duration of treatment was at least 8 weeks. Patients continued treatment until they have disease progression, become unable to tolerate or non-compliant with therapy, or withdraw consent.

4.2 MOLECULAR ANALYSES

1. Sample collection and storage

For PGX analyses, a 3mL whole blood sample was collected from each patient, preferably before the therapy starting, and stored in freezer at -80° C. Blood specimens were collected in sterile tubes with any anticoagulant agent but heparin was not admitted.

Collection, storage, and processing of blood specimens for PK of exemestane and estrogens level (estradiol (E_2) and estrone (E_1) and estrone sulfate(E_1S)) was carried out according to specific procedures. For the plasma PK analysis, blood (about 5 mL) was drawn into a pre-chilled heparinized tube and put in an ice-water bath until centrifuged. Samples had to be centrifuged at 4°C within 30 minutes of collection. The harvested plasma was transferred to two screw-cap polypropylene storage tubes. Plasma had to be stored frozen at approximately -20°C within 60 minutes of sample collection.

For serum estrogen determinations, 20 ml of venous whole blood was draw allowed to clot for at least 30 minutes, samples were then centrifuged (1200 g for 10 minutes at room temperature). The serum was transferred into four screw-cap polypropylene tubes (two aliquots for free estrogens and two aliquots for E_1S quantification) and stored at -20°C.

Blood sampling was performed at the start of exemestane treatment right before the intake of the daily exemestane dose and after 8 weeks of treatment. This sampling allowed the determination of the $C_{through}$ of exemestane, the basal levels of estrogens as well as the level of the estrogens suppression after 8 weeks of treatment.

All personal and clinical data were catalogued in appropriate databases, prepared in accordance with the Privacy Policy.

2. Genomic DNA extraction

The extraction of genomic DNA from whole blood was performed with the automated extractor BioRobot EZ1 (Qiagen SPA, Milano, Italy) (Fig. 6).

Figure 6. BioRobot® EZ1 for automated genomic DNA extraction from whole blood

The Card “EZ1 DNA Blood”, in association with the Kit “EZ1 DNA Blood Kit 350 µl”, was used for the extraction of genomic DNA from 350 µl of whole blood obtaining 200 µl as final volume, corresponding approximately to 5-12 µg of DNA. Once introduced the appropriate card and start the program, the BioRobot allows to process 6 samples simultaneously, without any intervention by the operator.
This technology is based on the use of silica-coated magnetic particles. DNA is isolated from lysates in one step through its binding to the silica surface of the particles in the presence of a chaotropic salt. The particles are then separated from the lysates using a magnet and the DNA is efficiently washed and eluted in elution buffer. In this way the DNA is held and purified from the blood sample. (Fig. 7)
DNA extracted is maintained at 2-8°C.

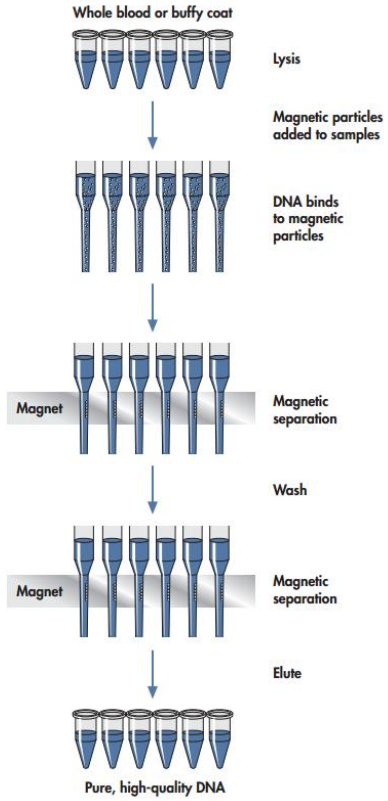


Figure 7. Principle of DNA extraction with EZ1 BioRobot

3. Polymerase chain reaction (PCR)

The technique of DNA polymerase chain reaction (PCR) was invented in 1983 by K. Mullis and allows to produce a large number of copies of a specific DNA sequence *in vitro*. It also allows to isolate and amplify any gene from any organism and then analyze the sequence, to perform cloning or mutagenesis procedures, or even to establish diagnostic tests that detect the presence of mutated forms of the gene. In the *in vitro* process, DNA is initially heated to temperatures close to boiling, in order to denature it and thus obtain single-stranded mold, then Taq polymerase is used to catalyze the duplication of the

parental strand (Fig. 8). To start the synthesis reaction, this enzyme requires a primer represented by a small sequence of double-stranded DNA.

In the reaction tube two primers are added, one to allow the synthesis of the sense strand (sense or forward primer) and one for the synthesis of the antisense (antisense or reverse primer). The two primers define the target region to be amplified (Fig. 9).

Figure 8. Thermal Cycler.

PCR requires several reagents and reaction conditions that vary with the time. In particular, the samples are subjected to a series of thermal cycles which are summarized below:

- An initial period at elevated temperature (94-95°C) that allows the DNA denaturation, in order to completely separate the template's strands that act as a mold.
- A variable number of consecutive cycles of amplification, each of which consists of three phases corresponding to three different temperatures:
    1. Complete DNA denaturation, carried out by heating at elevated temperature (94-95° C);
    2. Pairing (annealing) of sense and antisense primers with complementary sequences on the DNA template. In this phase the temperature is lowered to values which may vary from 50° C to 65° C according to the specific characteristics of the primers used;
    3. extension (elongation) of the primers and synthesis of new strands by the Taq polymerase, at a temperature of 72° C optimum for the enzyme activity.

    To obtain the amplification of the desired DNA sequence, this cycle of three steps must be repeated several times, typically from 25 to 40 times.
- A final period at 72° C to complete the elongation.

This series of thermal cycles is carried out thanks to a programmable instrument, the thermal cycler, capable of changing the temperature very quickly and keep it constant for a given period of time. The result of a PCR is that, at the end of *n* cycles of amplification, the reaction mixture contains a theoretical maximum number of double-stranded DNA equal to 2^n (where "*n*" represents the number of amplification cycles). In the first cycle of PCR the two primers anneal with the two strands of the

denatured template, thus providing the trigger for the polymerase that synthesizes complementary strands; as result of this cycle, two new strands, longer than the region to be amplified, whose end parts correspond to the sequence of the primers used to identify the target sequence, are created. In the second cycle, the primers anneal to the original template again and so it produces other new strands of undefined length. In subsequent cycles only fragments of the desired length are formed and they contain the specific region you want to amplify.

The starting material of the PCR is the genomic DNA containing the sequence to be amplified, it is not necessary to isolate this sequence because it is directly bounded by two specific primers used in the reaction. The reagents used in a PCR are: reaction buffer, magnesium ions supplied by the magnesium chloride ($MgCl_2$), deoxynucleotides triphosphate (dNTPs), the specific primers, DNA polymerase and the template. In particular, for each sample, a reaction mixture, containing the reaction buffer, a solution of $MgCl_2$, the dNTPs, primers and DNA polymerase, is made before adding genomic DNA.

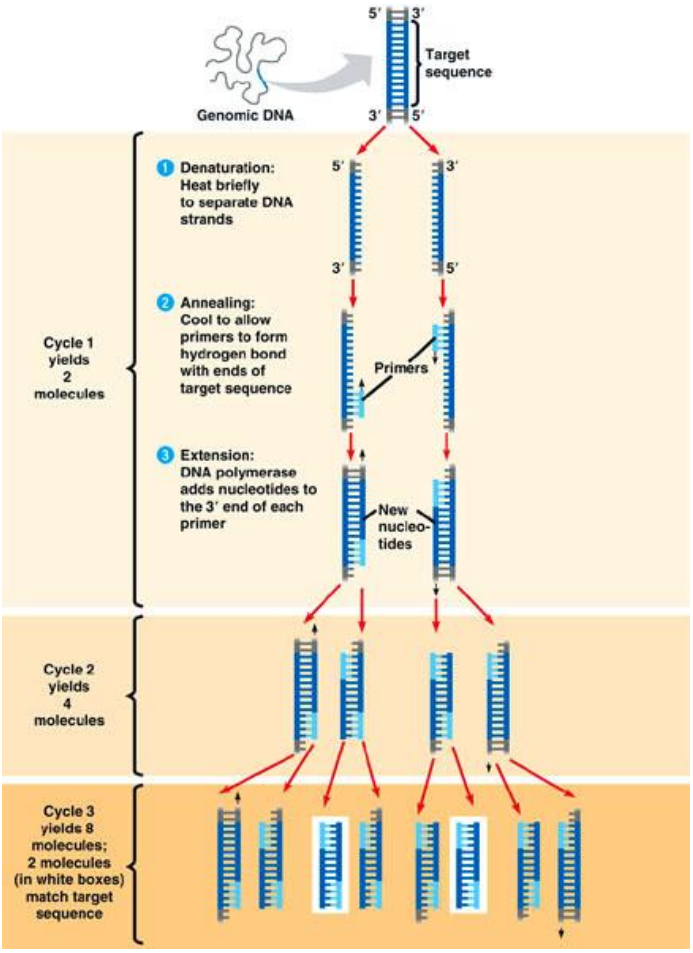


Figure 9. Polymerase Chain Reaction (PCR)

.

1) Reaction Buffer: it is a Tris-HCl and KCl based buffer and it is necessary to reproduce the optimal conditions for the activity of the polymerase thus increasing the throughput or the number of nucleotides that the enzyme can insert in succession before separating from the template strand.
2) Mg^{2+}: it is essential for the activity of Taq polymerase as its bond with the enzyme stabilizes it in a three-dimensional conformation that facilitates its activities. The Taq polymerase shows its highest activity around a concentration of free Mg^{2+} equal to 1.2-1.3 mM. Its concentration is, however, influenced by the concentration of nucleotides as there is a link between equimolar Mg^{2+} and dNTPs. You can then also use Mg^{2+} concentrations higher than those indicated above, but at higher concentrations polymerase tends to incorporate incorrect nucleotides.
3) dNTPs: the solutions of dNTPs contain the four nitrogenous bases of DNA: dATP, dGTP, dTTP and dCTP. For a good efficiency of the PCR the four nucleotides must be present in equimolar concentrations and the optimum concentration is around 50-200 µM. A too high concentration may increase the incorrect rate of incorporation, while a too low concentration may reduce the efficiency of the reaction.
4) Primers: primers design can be performed manually, or more frequently through the use of softwares that facilitate the choice such as "Primer3Plus" (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). The aim of primer design is to obtain a balance between two goals: efficiency and specificity of amplification. Given a target DNA sequence, primer design software attempts to strike a balance between these two goals by using pre-selected default values for each of the primer design available. In particular, optimal primer pairs should be closely matched in Melting Temperature (Tm) and must not be able to form loops and primer dimers. Primer length (about 20-base pairs), sequence and GC contents are taken into account to select proper primers sequences.
5) DNA polymerase: it is derived from the *Thermophilus acquaticus* bacterium, it is stable at high temperatures and works with maximum efficiency between 72°-75°C. The thermal stability is a critical feature of this enzyme. Taq polymerase at 72°C has an enzymatic activity that allows the incorporation of 50-60 nucleotides per second which corresponds to approximately 3 Kb per minute. The optimal concentration of Taq DNA polymerase ranges from 0.5 to 2.5 U. A too high concentration may decrease the specificity of the reaction, while a too low concentration may not enable the conclusion of all cycles.

Optimization of the PCR conditions

In order to obtain good results in the PCR process the three key parameters to consider are: efficiency (or yield), reaction specificity and accuracy. To perform a PCR with high efficiency, we have to obtain a large number of products with the least number of cycles as possible. The specificity of the reaction is

intended as the ability to amplify only the sequence of interest, without obtaining nonspecific products. Finally, a high accuracy is given by the presence of a negligible number of errors introduced by DNA polymerase.

Once the primers are designed, there are conditions that, if modified in an appropriate manner, can improve these parameters:

- ✓ Mg^{2+} concentration: the presence of divalent cations is critical, and it has been shown that magnesium ions are superior to manganese, and that calcium ions are ineffective. The optimal Mg^{2+} concentration for Taq polymerase efficiency must be set up to match dNTPs and primer concentration/ sequence. dNTPs are the major source of phosphate groups in the reaction, and any change in their concentration affects the concentration of available Mg^{2+}.
- ✓ Number of cycles**:** needs to be adjusted for each PCR protocol and it is set up checking the accumulation of target sequence by gel electrophoresis after each amplification reaction. The correct number of cycles to be used should guarantee a sufficient balance between efficiency and specificity of amplification. Thirty–eight cycles of amplification resulted appropriate for most of the protocols employed in this work.
- ✓ Annealing Temperature: it is set up performing a gradient temperature PCR (Mastercycler Gradient Eppendorf, Hamburg, Germany). Annealing temperatures could be calculated by several methods considering the Tm of the primer-template pairs. However, in practice, because the Tm is variously affected by the individual buffer components and the primer and template concentrations, any calculated Tm value should be considered just as a first approximation. A range of 60±5°C was tested by 1 degree increments to find the optimal reaction conditions.
- ✓ Additives: they may be added to the reaction mixture of PCR to increase the specificity of annealing of primers or the amount of amplified products. They are denaturing substances that lead to a destabilization between the bases and, consequently, also to a high destabilization of the complex primer/aspecific DNA. The substances that may be used are: dimethylsulfoxide (DMSO) (up to 10%), formamide (up to 5%) and glycerol (up to 10-15%).

4. Methodologies for polymorphisms analysis

NCBI (National Center for Biotechnology Information) databases and tools were used to select the analyzed polymorphisms. The NCBI presents a web site showing links to databases containing information about genes (Gene), polymorphisms (dbSNP), scientific literature (PubMed), besides search and analysis tools. These and other additional databases (SNP500, PharmGKB (The Pharmacogenomics Knowledge Base), and 1000 Genomes Browser) were consulted for assay design (genetic sequences, polymorphisms description, primer design), during this work of thesis.

Subsequently, according to the type of polymorphism to be analyzed and to the specific characteristics of the nucleotide sequence, the most suitable method of genotyping has been chosen.

In particular, in this PhD thesis, semi-automated, recently developed genotyping methods have been used. These are based on PCR reactions and allow the identification of genetic polymorphisms in a very simple and easy way: Pyrosequencing (PSQ), allelic discrimination based on TaqMan chemistry, the Fragment Analysis (Gene Scan).

In a second phase, analytical platforms were implemented with the introduction of Illumina BeadXpress ® Reader, based on GoldenGate chemistry and VeraCode Beads technology.

a. Pyrosequencing

Pyrosequencing is an analytical technology for SNP identification consisting of a real-time pyrophosphate detection method (Fakhrai-Rad *et al.*, 2002; Ronaghi *et al.*, 2001).

This technique is based on indirect bioluminometric assay of the pyrophosphate (PPi) that is released from each dNTP upon DNA chain elongation. Following Klenow polymerase mediated base incorporation, PPi is released and used as substrate, together with adenosine 5'-phosphosulfate, for the ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by luciferase. The ensuing light output is proportional to the number of added bases, up to about four bases. To allow processivity of the method, dNTPs in excess are degraded by apyrase, which is also present in the starting reaction mixture and continuously degrades ATP and unincorporated dNTPs. This switches off the light and regenerates the reaction solution. The dNTPs are added one by one to the template during sequencing procedure. It should be noted that deoxyadenosine alfa-thio triphosphate is used as a substitute for the natural dATP since it is efficiently used by the DNA polymerase, but not recognized by the luciferase. The process is fully automated and adapted to a 96-well format, which allows rapid screening of large panel of samples (Fig.10).

Following a first phase of sample preparation, the plate is loaded on an instrument, the PSQ 96MA Pyrosequencing (Fig. 11), which determines and provides directly the genotype at the level of the analyzed SNP.

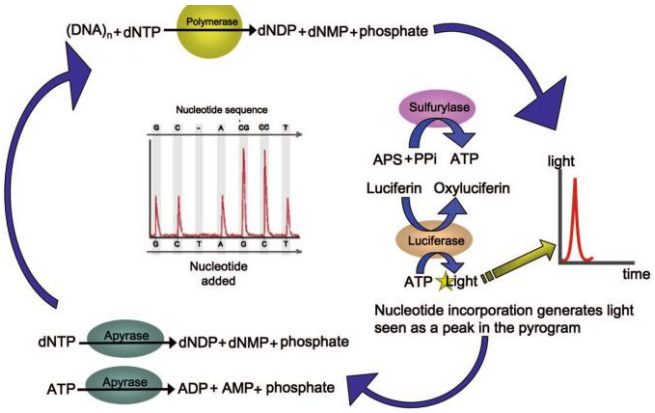


Figure 10. Pyrosequencing chemistry: biochemical reactions and enzymes involved in the generation of light signals by DNA pyrosequencing. Each peak in the pyrograms represents a pulse of light detected in the instrument. dNTP, deoxynucleoside triphosphate; dNDP, deoxynucleoside diphosphate; dNMP, deoxynucleoside monophosphate; PPi, pyrophosphate; APS, adenosine 5-phosphosulfate.

Figure 11. PSQ 96MA Pyrosequencing

Pyrosequencing analysis is performed on PCR-amplified DNA. One of the PCR primers must be biotin-labeled for immobilization to streptavidin coated Sepharose beads (Fig.12). This allows the separation of the two DNA strands produced by PCR, since the assay must be carried out on single stranded DNA. If reverse primer is biotynilated we have the forward assay, otherwise, if the forward primer is biotin-labeled, the assay is called reverse.

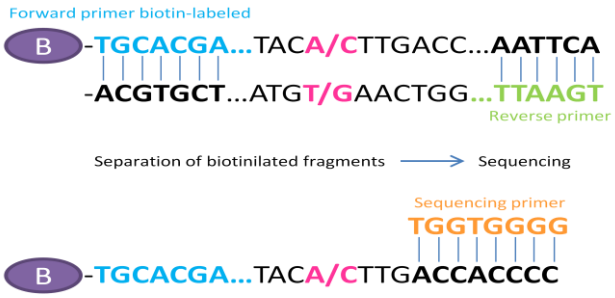


Figure 12. Schematic representation of Pyrosequencing method. SNP position is indicated by the slash (/). "B" stands for biotin.

PCR reaction product is mixed with streptavidin coated High Performance Sepharose beads (Amersham Biosciences, Uppsala, Sweden) in the presence of a binding buffer (Tris 10 mM, Sodium Chloride 2 M, EDTA 1 mM and Tween 20 0.1%, pH 7.6). The mixture is allowed to shake for 10 minutes at room temperature. The samples are subsequently transferred to a 96-well filter plate and vacuum (vacuum manifold for 96 well filter plate, Millipore) is applied to remove all liquid. Denaturation solution (Sodium Hydroxide 0.2 M) is added to denature double stranded PCR product DNA. After 1 minute incubation, vacuum is applied to remove the solution and the non immobilized DNA. The beads are washed twice with a washing buffer (Tris 10 mM, pH 7.6) in the presence of the vacuum. The beads with the immobilized template are resuspended by adding 45 µl annealing buffer (Tris 20 mM, Magnesium Acetate Tetra-Hydrate 2 mM, pH 7.6) and sequencing primer (2 µM) is added to each sample. Design of sequencing primers for Pyrosequencing follows the same criteria as for the PCR primers, except that the Tm of these primer may, if necessary, be lowered. The sequencing primer could thus be shorter than the PCR primers, typically 15 bp. The position of the primer is flexible within 5 bases from the SNP and can be designed on both the positive (reverse assay) or on the negative (forward assay) strand. Thirty-five µl of this mixture is transferred to a Pyrosequencing 96 wells plate (PSQ 96 Plate Low).

The plate is incubated for 5 minutes at 60° C to allow complete sequencing primer annealing on the template DNA. After samples cooling, the plate is transferred on the Pyrosequencing instrument. The biotin labeled DNA template, annealed to the sequencing primer, is incubated with enzymes (DNA polymerase, ATP sulfurylase, luciferase and apyrase) and the substrates (adenosine 5'phosphosulfate and luciferin). The first of four dNTPs is added to the reaction. DNA polymerase catalyzes the incorporation of the dNTP into the DNA strand, complementary to the base in the template strand. Each incorporation event is followed by the previously described reaction cascade, leading to the generation of visible light in amount that is proportional to the number of nucleotide added. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and seen as peak in a pyrogram. The height of each peak (light signal) is proportional to the number of incorporated nucleotides (Fig. 12).

As the process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the pyrogram.

"PSQ Assay Design" software was used for the planning of the described assays: it allows to easily choose the set of primers (sense and antisense primers for PCR and sequencing primer for subsequent analysis at PSQ) most suitable for the study of each SNP The analysis of the results is accomplished with the"PSQ^{TM} 96 MA software".

The reagents and solutions used in Pyrosequencing analysis are:

- Aqua B. Braun Ecotainer, sterile water for injection (B. Braun, Melsugen AG, Germany);
- Streptavidin Sepharose^{TM} High Performance (Amersham Biosciences AB, Uppsala, Sweden);

- Sequencing primer provided in lyophilized form (Sigma Genosys, Cambridge, UK) and then resuspended in sterile water to obtain a final concentration of 100 μM;
- Pyro Gold Reagents Kit (Biotage AB, Uppsala, Sweden) constituted by:
  - Enzyme mixture (luciferase, DNA polymerase, apyrase, sulfurylase and proteins binding to single-stranded DNA- provided in lyophilized form and then resuspended in sterile water;
  - Substrate mixture (adenosina 5'fosfosulfato [APS] and luciferin) provided in lyophilized form and then resuspended in sterile water;
  - dATPαS in solution;
  - dCTP in solution;
  - dGTP in solution;
  - dTTP in solution.
- PSQTM 96 Sample Preparation Kit (Biotage AB, Uppsala, Sweden) constituted by:
  - Binding Buffer (10 mM Tris-HCl, 2M NaCl, 1 mM EDTA, 0,1% Tween 20; pH=7,6);
  - Denaturation Solution (0,2 M NaOH);
  - Washing Buffer (10 mM Tris-acetate; pH=7,6);
  - Annealing Buffer (20 mM Tris-acetate, 2 mM Mg^{2+}-acetate; pH=7,6).

Other materials and instruments used in Pyrosequencing methodology are:

- PSQ 96 Plate Low (Biotage AB, Uppsala, Sweden);
- 96-well filter plates (Millipore, MA, USA);
- PSQTM 96 Reagent Cartridge (Biotage AB, Uppsala, Sweden);
- PSQTM 96 Sample Prep Tool Termoplate (Biotage AB, Uppsala, Sweden);
- Multichannel Pipette (Matrix Technologies Corporations, NH, USA);
- Vacuum pump (Millipore, MA, USA);
- Shaker (Analitica De Mori, MI, Italia);
- PyroMarkTM Vacuum Prep Workstation (Biotage AB, Uppsala, Sweden);
- PSQ96 MA Pyrosequencing (Biotage AB, Uppsala, Sweden), software PSQTM 96 MA;
- Pyrosequencing Assay Design Software, version 1.0.6 (Biotage, Westbrough, MA, USA).

The SNPs analyzed with Pyrosequencing technology (PSQ) are reported in Table 4:

GENE	SNP NAME	Rs ID	FUNCTION	AA CHANGE	SEQUENCE
CYP19A1	47T/C	rs700519	Missense	Arg264Cys	AGAAGTTCTGATAGCAGAAAAAAGA[C/T]GCAGGATTTCCACAGAAGAGAAAC
CYP3A4	*1B_-392A/G	rs2740574	5'UTR	NA	GAGGACAGCCATAGAGACAAGGGCA[A/G]GAGAGAGGCGATTTAATAGATTTT
RIZ1	delP704	rs2308040	Deletion	delPro704	TCAGTTGCTGAAATCCCTGCAGGAG[-/GAG]TTAGTTTATCTTGGGTTTGAAGAA

Table 4. SNPs analyzed by Pyrosequencing technology.

b. TaqMan® Assay

The allelic discrimination consists in the determination of the two variants of a single nucleic acid sequence by means of the "5' fluorogenic nuclease assay". In particular, this technique exploits the exonuclease property in the direction 5'→3' of the Taq polymerase when it encounters, during its activity of DNA fragment elongation, an oligonucleotide perfectly matched with the DNA template employed for the elongation. Through this method it is possible to investigate SNPs.

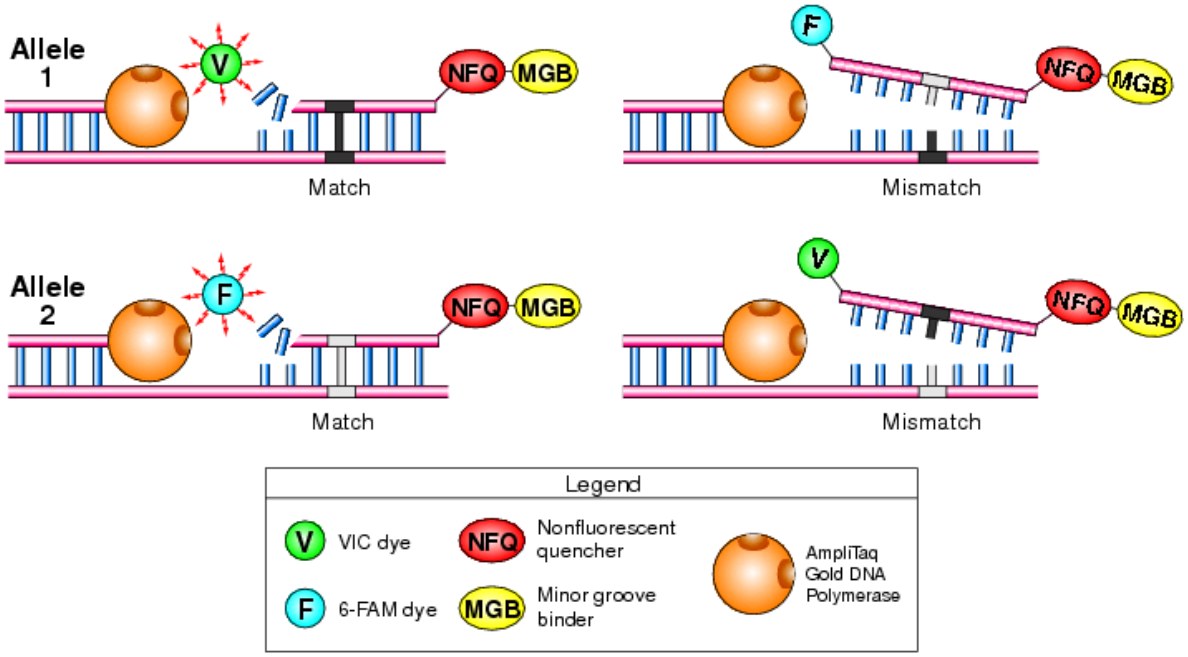


Figure 13. Schematic representation of TaqMan® technology.

TaqMan® allelic discrimination is based on the use of a Real Time PCR (RT PCR), that, in addition to the sense and antisense primers needed for the amplification of the SNP containing fragment, involves the use of an oligonucleotide *(probe)* able to pairing with the template. The probe pairs in an intermediate position between the sense and the antisense primer. The probe is functionalized at the two ends: in one part there is a "quencher" fluorophore (TAMRA) which acts as a silencer of fluorescence, the other one is tied to a "reporter" constituted by a fluorescent fluorophore (FAM or VIC). The action of silencing by the quencher occurs by transfer of energy from one fluorochrome to the other one when they are near to each other. In the reaction two different allele-specific probes, labeled with different fluorophores (fluorochrome FAM or VIC), are placed: one contains a perfect match to the wild type (allele 1) and the other one presents a perfect match to the mutation (allele 2). The allelic discrimination assay classifies unknown samples as: homozygous and heterozygous.

TaqMan probe-based chemistry uses a fluorogenic probe to detect specific PCR product as it accumulates during PCR cycles. In Fig. 13 is reported a schematic representation of a TaqMan Assay. During the denaturation step, the reporter (R) and the quencher (Q) are attached to the 5' and 3' ends

of a TaqMan probe. When both dyes are attached to the probe, reporter dye emission is quenched. During each extension cycle, the hot-start DNA polymerase system cleaves the reporter dye from the probe. After being separated from the quencher, the reporter dye emits its characteristic fluorescence which is recorded by a detector.

The probes are chosen according to certain characteristics:

- The T_m must be at least 5° C higher than the T_m of the two PCR primers because they must bind to the nucleotide sequence when executing the synthesis of the complementary strand;
- The oligonucleotide must have a length of about 20-30 bp and 50% of G and C;
- The extension phase must be performed at a temperature lower than the 72°C used in the PCR, in order not to cause the detachment of the probe from the template (for this reason we use high concentrations of $MgCl_2$);
- The probe must not form dimers or even pair with itself.

Figure 14. Applied Biosystems 7500 Real-Time PCR System instrument.

Samples are analyzed using the Applied Biosystems 7500 Real-Time PCR System instrument (Fig. 14). The allelic discrimination was performed with the SDS software 2.3 (Applied Biosystems).

For SNP assay a preformed “TaqMan® SNP Genotyping Assay” is employed: it is available on-line in the catalogue of Applied Biosystems (http://www3.appliedbiosystems.com/AB_Home/index.htm). As an alternative, you can use the service offered by the same company that, on sending the gene sequence containing the nucleotide variation, develops and tests specifically an assay called "Custom SNP Genotyping assay TaqMan®".

The practical procedure of the TaqMan® technology is really very simple and allows to analyze quickly the genotype using a universal mix (master mix) and a solution containing PCR primers and the two allele-specific probes. The step of sample preparation involves the use of 96-well plates with specific optical properties. The reaction mixture is prepared by combining the specific mix for the SNP under investigation (SNP Assay 20X or 40X), containing primers (sense and antisense) and the two probes

labeled with FAM or VIC, to the Master Mix (TaqMan Genotyping Master Mix 2X) universal for all genotyping analyses, containing dNTPs, Taq Polymerase, $MgCl_2$ and salts in a suitable concentration creating an adequately buffered environment. The solution is dispensed into wells and, finally, genomic DNA is added (approximately 20 ng of DNA for each sample).

Once set up, the plate is covered with an adhesive film and centrifuged for a few minutes to eliminate the presence of any air bubbles at the bottom of the wells. Then the plate is loaded into the ABI PRISM 7900HT machine, at this stage the RT-PCR conditions (temperature, duration and cycles) and the test volumes (20 µl) are determined, and the markers (FAM and VIC) are assigned to polymorphism's alleles. The amplification is carried out with a thermal cycler integrated into the instrument using the following thermal profile:

- 50° C for 2 minutes;
- 95° C for 10 minutes;
- 40 cycles for (92° C for 15 seconds; 60° C for 1 minute)

At the end of the PCR reaction an end point scanning of the 96-well plate containing the samples is carried out, in order to detect the fluorescence signal produced in each well by the two fluorophores (FAM and VIC) associated to the allele-specific probes. Finally, thanks to the processing of obtained data by software SDS 2.3, the assignment of the genotype corresponding to each sample occurs.

For the analysis with TaqMan® technology were used the following reagents:

- 2X TaqMan Genotyping Master Mix (Applied Biosystems, CA, USA);
- 20X or 40X "TaqMan® SNP Genotyping Assay" or "CustomTaqMan® SNP Genotyping Assay"(Applied Biosystems, CA, USA);
- MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems, CA, USA);
- Optical Adhesive Covers (Applied Biosystems, CA, USA);
- Real-Time ABI PRISM 7900HT instrument (Applied Biosystems, CA, USA);
- SDS 2.3 software (Applied Biosystems, CA, USA).

The SNPs analyzed with the TaqMan® Assay are reported in Table 5:

GENE	SNP NAME	Rs ID	FUNCTION	AA CHANGE	SEQUENCE
CYP19A1	Ex11+410A/C	rs4646	3'UTR	NA	CTCTGGTGTGAACAGGAGCAGATGAC[A/C]AATAGCACCTAGCTTGGTGACAAC
CYP19A1	1558T/C	rs10046	3'UTR	NA	AACACTAGAGAAGGCTGGTCAGTACC[C/T]ACTCTGGAGCATTTCTCATCAGTAG
CYP17A1	27A/G	rs743572	5'UTR	NA	GGGTGCCGGCAGGCAAGATAGACAGC[A/G]GTGGAGTAGAAGAGCTGTGGCAA
CYP1B1	*3_4326G/C	rs1056836	Missense	Leu432Val	ACCAGTGGTCTGTGAATCATGACCCA[C/G]TGAAGTGGCCTAACCCGGAGAACTT
CYP3A5	*3_6986A/G	rs776746	Intron 3	NA	TCTCTTTAAAGAGCTCTTTTGTCTTTCA[A/G]TATCTCTTCCCTGTTTGGACCACATT
COMT	12A/G	rs4680	Missense	Val158Met	CCCAGCGGATGGTGGATTTCGCTGGC[A/G]TGAAGGACAAGGTGTGCATGCCTG
ESR1	497T/C	rs2234693	Intron	NA	TTCATCTGAGTTCCAAATGTCCCAGC[C/T]GTTTTATGCTTTGTCTCTGTTTCCC
ESR1	256A/G	rs9340799	Intron	NA	TTTCCCAGAGACCCTGAGTGTGGTCT[A/G]GAGTTGGGATGAGCATTGGTCTCTA
ESR2	1082A/G	rs1256049	Cds-synon	Val328Val	GTGGAGCTCAGCCTGTTCGACCAAGT[A/G]CGGCTCTTGGAGAGCTGTTGGATG
ESR2	1730A/G	rs4986938	3'UTR	NA	AGGTGAACTGGCCCACAGAGGTCACA[A/G]GCTGAAGCGTGAACTCCAGTGTGT

Table 5. SNPs analyzed by TaqMan® method.

c. Automated Fragment Analysis

Automated fragment analysis is performed to detect small variation in the length of a DNA fragment. It is based on the capillary electrophoresis coupled with fluorescence detection. Capillary electrophoresis occurs when an electric field is applied to an electrolyte solution within a capillary, causing ions migration. DNA fragments, having a negative charge, move toward the anode (+) and are separated by size (Fig. 15).

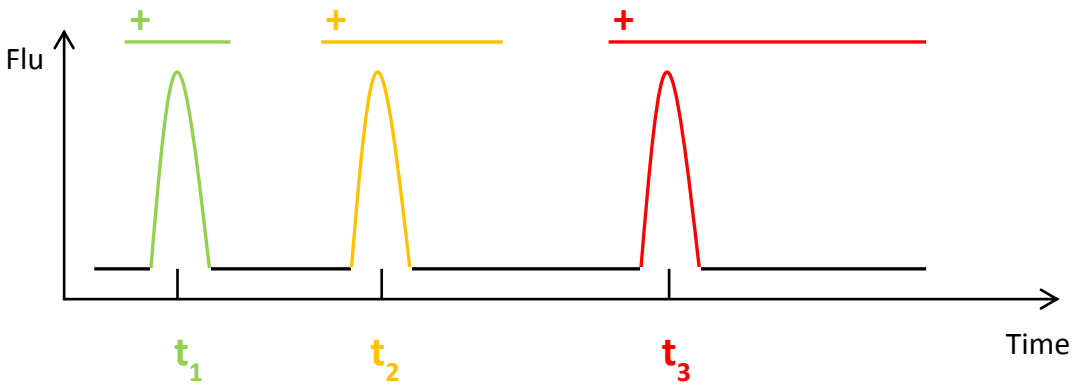


Figure 15. Scheme of the separation of different sized DNA fragments labeled with different fluorophores (ROX=red, JOE=green, LIZ=yellow) by capillary gel electrophoresis.

DNA fragments are produced by PCR employing one 5' fluorescence labeled primer with the HEX (isomer-free succinimidyl ester of 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein, excitation and emission maxima of 535 and 556 nm respectively) fluorophore. In the analysis a marker of DNA molecular weight labeled with a different fluorophore, the ROX, is also employed. It serves as internal standard (Fig. 16).

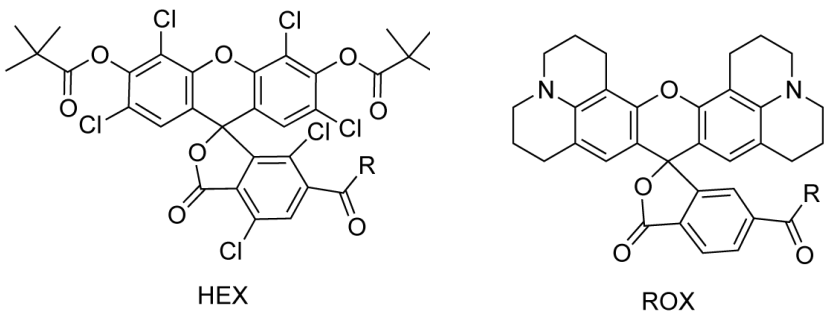


Figure 16. Molecular structure of the two fluorophore used to characterize the PCR sense primer (HEX) and the molecular marker (ROX) for automated fragment analysis.

These dye-labeled fragments are detected by fluorescence and in turn rendered into a sequence or sized fragment. The pherogram analyzed by the software presents on the abscissa the separated molecular weight fragments, while on the ordinate the intensity of the fluorescence peak. The samples are

analyzed in the Genetic Analyzer ABI Prism 3100 instrument (Applied Biosystems) (Fig. 17). Gene Scan analysis software (Applied Biosystems) allows data extraction and elaboration.

Figure 17. Genetic Analyzer ABI Prism 3100.

The process is very simple. The first phase consists in the amplification of the gene fragment containing the polymorphism of interest by means of a PCR that presents one of the two primers labeled with the fluorophore HEX covalently linked in 5' (not reactive extremity).

Since this method is very sensitive, a small concentration of amplified fragment to conduct the analysis is sufficient. Consequently, the samples, after being analyzed by electrophoresis on agarose gel, are suitably diluted. The mix needed to perform the analysis consists of 14.5 µl of deionized and purified formamide and 0.5 µl of Internal Lane Size Standard [ROX] for each sample. Once prepared the mix, this is aliquoted into a 96-well plate and, subsequently, 1 µl of the diluted sample is added to obtain a total of 16 µl per well. The formamide is a strong denaturant and is sufficient the contact with the DNA to exert its effect. The plate is covered, to prevent evaporation of the solution, and denaturated (2 min at 95° C). Immediately after denaturation, the plate is placed in ice to avoid the rewinding of DNA strands and it is loaded into the Genetic Analyzer ABI PRISM 3100 instrument (Fig. 18).

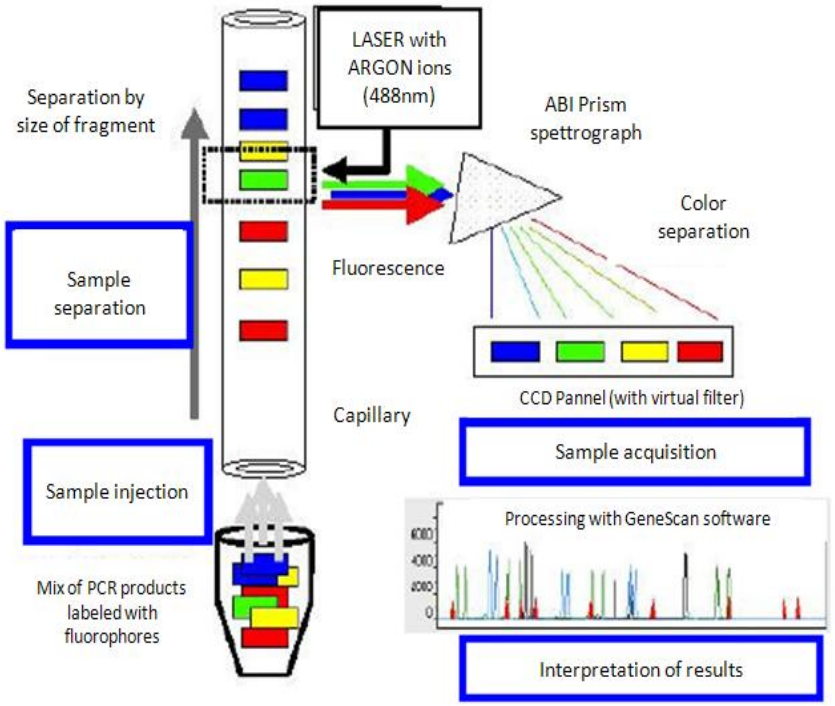


Figure 18. Scheme of the processing of DNA fragments within the Genetic Analyzer ABI PRISM 3100 instrument. The involved steps are: injection of the samples, separation of the fragments by size using electrophoresis, determination of fluorescence with CCD camera, interpretation of the results with the Gene Scan software.

Once started the instrument, the 16 capillaries penetrate in the plate and take samples. The loading of the samples takes place through electrokinetic injection, i.e. through the application of a potential of 15 KV for about 5 seconds which moves all the charged molecules within the capillary. There are activities of competition by charged molecules or ions, present in the sample, which can interfere with this delicate phase of the process. The sample dilution in sterile water and purified formamide is also useful to reduce these interference phenomena. The phenomenon of stacking , which allows to the fragments and the mix to be loaded into a restricted and compact zone of the capillary, ensures the correct injection of the samples in the capillaries avoiding DNA diffusion phenomena. Stacking permits to produce an area of low conductivity, and this is made possible from immersion of the capillary in water before loading the samples. After the first phase of injection, the samples are separated by an electrophoretic run and, at the exit of the capillary, they are bombarded by a laser that excites all fluorophores emitting fluorescence in different regions of the spectrum. An analyzer of multiple wavelengths CCD camera (charged-coupled device) identifies the emissions of each fragment passing through the detector. The analysis of fluorescence occurs both for unknowns fragments and for the internal standard fragments (Fig. 19).

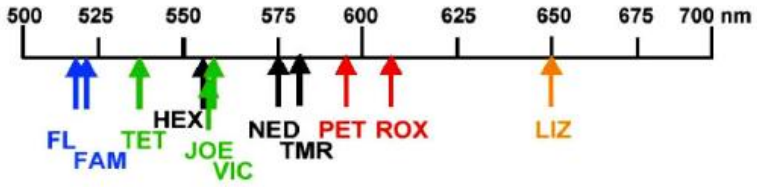


Figure 19. Range of the visible spectrum captured in CCD camera

The ABI 3100 data collection software allows to control the conditions of electrophoresis and manages the creation of samples files and lists of injections. The extraction and processing of data are managed by the Gene Scan analysis software that allows the conversion of the data into appropriate colored peaks which have assigned values of fragment length, based on the time of output and the type of emission. The instrument Genetic Analyzer ABI PRISM 3100 (Applied Biosystems, Foster City, CA) was also employed, managed by the Gene Scan analysis software (Applied Biosystems, Foster City, CA).

The success of an electrophoretic run depends on several factors (Butler *et al.*, 2004):

- The capillary: in capillary electrophoresis the separation is performed using a tube with an internal diameter of 50-100 micron. This thinness allows the application of a high electric field and therefore reduces the times of escalation without overheating problems associated with the high voltage used.
- The polymer: there are many different types of means of sieving used in electrophoretic separations. For example physical gels, products from agarose, or the common chemical gels, such as polyacrylamide used in electrophoretic denaturants gel plates, which are cross-linked rigid materials in which the porous structure is linked through covalent bonds, could be used. These two materials are however problematic to be used inside a capillary for the formation, for example, of air bubbles, both during the coating of the lumen and in the phase of gel contraction, due to the polymerization. The third type of sieving materials are tangled polymers also characterized by intermolecular interactions. These (for example the linear polyacrylamide that is not cross-linked) have replaced the other two types of gels as they are less problematic. For automated fragment analyses conducted for this thesis POP7 was used.
- The formamide: the use high-quality formamide with low conductivity is very important. Such substance in fact generates ionic products from its decomposition, including formic acid, that is negatively charged at neutral pH and would compete with DNA in the run on the capillary. This can also cause problems of both sensitivity and resolution.
- The buffer: the solution used to dissolve the polymers is important to stabilize and solubilize the DNA, provide charge carriers to the electrophoretic current and to increase the injection. If the concentration and the concomitant conductivity of the buffer are too high, the column will overheat and as a result will lose resolution.

- The temperature: to maintain the DNA denaturation obtained during sample preparationas previously described, the temperature of the column must be higher than room temperature. Furthermore, the internal standard is sensitive to temperature variations and, therefore, it can also be used as an indicator of a stable and well calibrated system.

The reagents and solutions used for this methodology are:

- Aqua B. Braun Ecotainer, sterile water for injection (B. Braun, Melsugen AG, Germany);
- Hi-Di™ Formamide (Applied Biosystems, Foster City, CA);
- Fluorophore ROX™ DYE (5-carboxy-X-rhodamine, succinimidyl ester) Gene Scan™ 400HD [ROX] Size Standard, (Applied Biosystems, Foster City, CA).

The Short Tandem Repeats (STRs) polymorphisms analyzed with the fragment analysis method are reported in Table 6:

GENE	STR NAME	Rs ID	FUNCTION	AA CHANGE	SEQUENCE
CYP19A1	(TTTA)n	rs60271534	Intron 4	NA	ATCATATTTTTAAAATAT[(TTTA)7/8/9/10/11/12/13]TTGAGACAGGCTCTGACTC
UGT1A1	*28_TA(6/7)	rs8175347	5'UTR	NA	CTTGGTGTATCGATTGGTTTTTGCCA[(TA)6/7]AGTAGGAGAGGGCGAACCTCTGG

Table 6. STRs analyzed by automated fragment analysis

d. Beadxpress reader coupled with VeraCode® technology and GoldenGate® Assay

Illumina BeadXpress Reader (Illumina, La Jolla, CA) is a dual-color laser scanning system allowing users to analyze several genetic markers in a multiplexing manner exploiting the VeraCode™ microbeads digital technology. This technology allows several types of multiplex testing ranging from genotyping, gene expression, RNA and protein-based assays, methylation and expression studies of 1 to 384 biomarkers per well at the same time. The VeraCode™ system is based on the VeraCode Beads, glass microcylinders (240 µm in length by 28 µm in diameter), each inscribed with a unique digital holographic code to unambiguously designate and track the specific analyte or genotype of interest throughout the multiplex reaction.

Unlike traditional microarrays, the VeraCode microbeads are used in solution, which takes advantage of solution-phase kinetics for more rapid hybridization times.

The microbeads highly pure glass, stable at high temperatures and chemical agents, represents an optimal surface for biomolecules attachment. In the GoldenGate Genotyping® Assay, each microbead is functionalized with a specific oligonucleotide which univocally identifies a single SNP. VeraCode™ beads are used for analyzing up to 384 genetic markers per sample in plates containing 96 samples each.

To perform a BeadXpress analysis, a sample preparation phase is required. In this phase, a unique multiple PCR reaction is performed. Subsequently, each SNP-containing fragment produced, is conjugated with a specific VeraCode microbead for the genotyping attribution analyses.

The first step in the GoldenGate Assay is DNA activation by biotinylation, which enables genomic DNA samples to bind (by biotin-streptavidin interaction) to paramagnetic particles (Fig. 20). This activation process is highly robust and requires only 250 ng of genomic DNA.

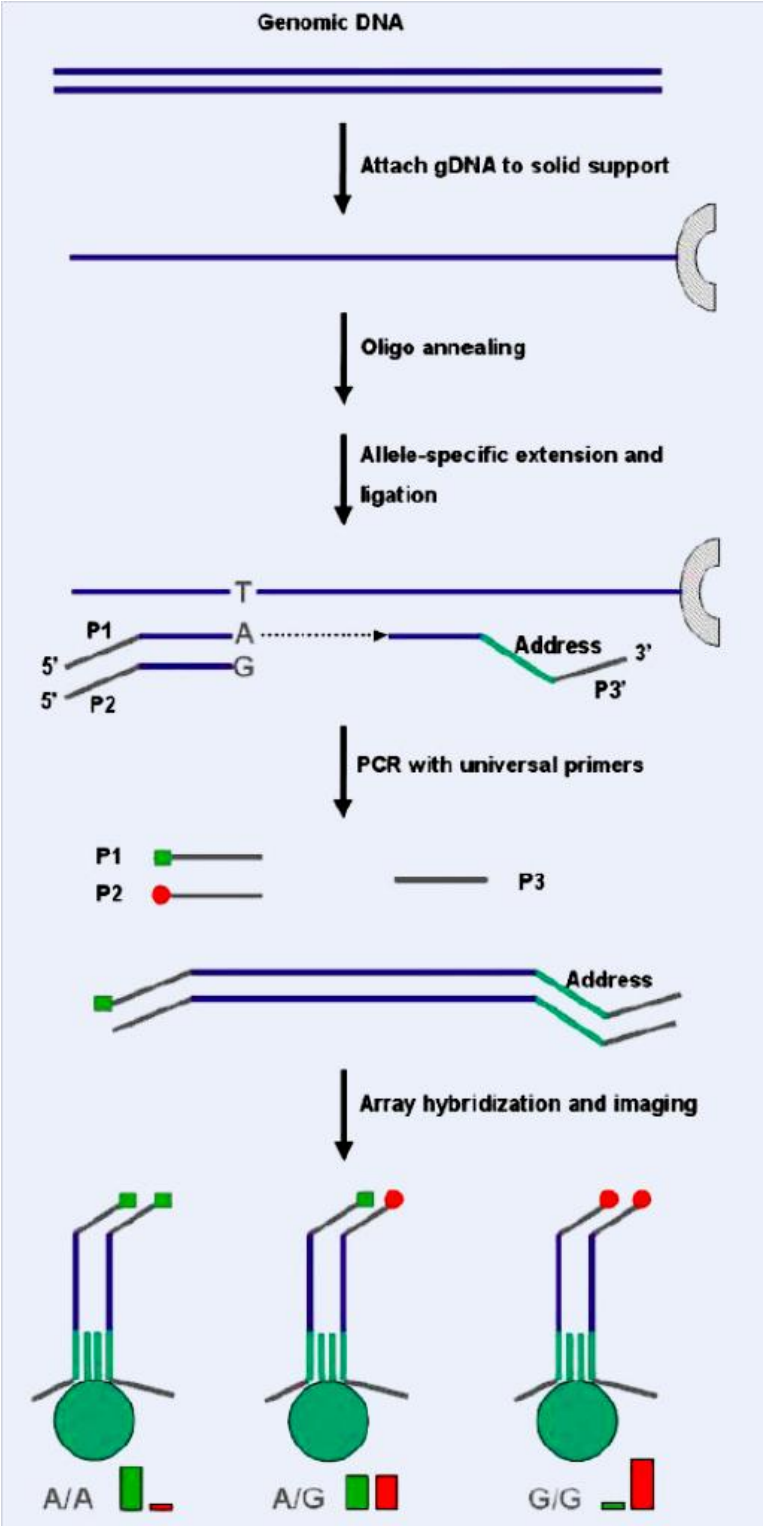


Figure 20. A schematic view of the GoldenGate Assay process.

Three oligonucleotides are designed for each SNP locus. For each SNP site there are two allele-specific oligos (ASOs). They have exactly the same sequence but differ only for the last nucleotide, which matches the polymorphic one found at the SNP site in the sequence of interest. So, for each DNA strand, only one ASO hybridizes, depending on the SNP genotype. A third oligo, the locus-specific oligo (LSO), instead, hybridizes several bases downstream from the SNP site. All three oligonucleotide sequences contain universal PCR primer sites (that is, complementary sequences recognized by the universal PCR primers P1, P2, P3); the LSO contains a unique address sequence (or "Lumicode") that targets a particular oligonucleotide-coated VeraCode microbead type. Assay oligonucleotides (ASOs and LSOs), hybridization buffer, and paramagnetic particles are then combined with the activated DNA in the hybridization step. During the primer hybridization process, ASOs and LSOs hybridize to the genomic DNA sample bound to paramagnetic particles. Because hybridization occurs prior to any amplification steps, no amplification bias is introduced into the assay. Following hybridization, several wash steps are performed, removing excess and mis-hybridized oligonucleotides. Extension of the appropriate ASO and ligation of the extended product to the LSO join information about the genotype present at the SNP site to the address sequence on the LSO. The ligation products (containing the SNP and the address sequence) serve as the PCR templates for universal PCR primers P1, P2, and P3. Primers P1 and P2 are fluorophore-labeled with Cy3- and Cy5-dyes, respectively, so, depending on the allele, the instrument will detect a one color (in case of homozygosis) or a two colors fluorescence (in case of heterozygosis). P3 primer is the only reverse primer at the locus specific site, allowing the amplification of the address sequence for the binding with a specific bead. After downstream processing, the single-stranded, dye-labeled PCR products are hybridized to their complementary bead type through their unique address sequences. Hybridization of the GoldenGate Assay products onto the VeraCode beads separates the assay products for individual SNP genotype readout.

After hybridization, the BeadXpress® Reader is used for microbead code identification and fluorescent signal detection.

The plate is loaded in the BeadXpress Reader and beads from 8 wells at a time are drawn up and aspirated onto the 8-chambered transparent groove plate in which, thanks to a combination of fluid flow, gravity and capillary force, they populate and align closely within the grooves. Once the beads are aligned, the entire fluidic cell is actuated across the optical system and scanned for fluorescent intensity and code classification. Here, a dual-color laser detection system identifies on one hand the unique holographic code embedded in each VeraCode bead and on the other hand the signal intensity associated with each bead discriminating the genotype. Assays developed with VeraCode microbeads typically include up to 30 replicates of each bead type. Each microbead is optically scanned up to a dozen times providing about 300 independent data point for each analyte ensuring reliable and accurate results.

The plate preparation process lasts about two days and the workflow is summarized in Table 7.

Process	Time	Day
DNA activation	1h 20 min	1
Oligonucleotides addition and DNA binding to paramagnetic particles	3h	1
Oligonucleotides-DNA binding	50 min	1
PCR - amplification	2h 30 min	1
Amplicons isolation	1h e 40 min	2
Amplicons hybridization with *Veracode Bead*	3h	2
Veracode Bead Plate washing	10 min	2
Veracode Bead Plate reading	1h 10 min-96 polymorphisms 3h 30 min-384 polymorphisms	2

Table 7. BeadXpress workflow.

Data generated using the BeadXpress Reader can be analyzed with Illumina's GenomeStudio™ data analysis software, which performs automated genotype clustering and calling.

The software permits the association between the fluorescence data and the correspondent genotype. A clusterization algorithm assembles in three groups the fluorescence values related to each sample based on the presence of only one (in case of homozygous genotypes) or two (heterozygous genotype) fluorescence signals. This process lets the software call the genotypes for each SNP investigated. The holographic code links the genotype call to a specific sample.

The graphical display of genotypes in GenomeStudio is a Genoplot, with data points color-coded for the call (red = AA genotype, purple = AB genotype, blue = BB genotype).

Genotypes are called for each sample with a dot by two coordinates representing their signal intensity (norm R) on the y axis and Allele Frequency (Norm Theta) on the x axis relative to canonical cluster positions (dark shading) for a given SNP marker.

Genome Studio normalizes the intensity of each fluorescence, so the Theta angle between the sample dot and the x axis is converted into a value on the x axis (Norm Theta), while in the y axis the fluorescence intensity is reported as a Theta angle normalized value (Norm R), approximately ranging from 0 to 1 (with some exceptions for outliers which can reach values of 2 or more) (Fig. 21).

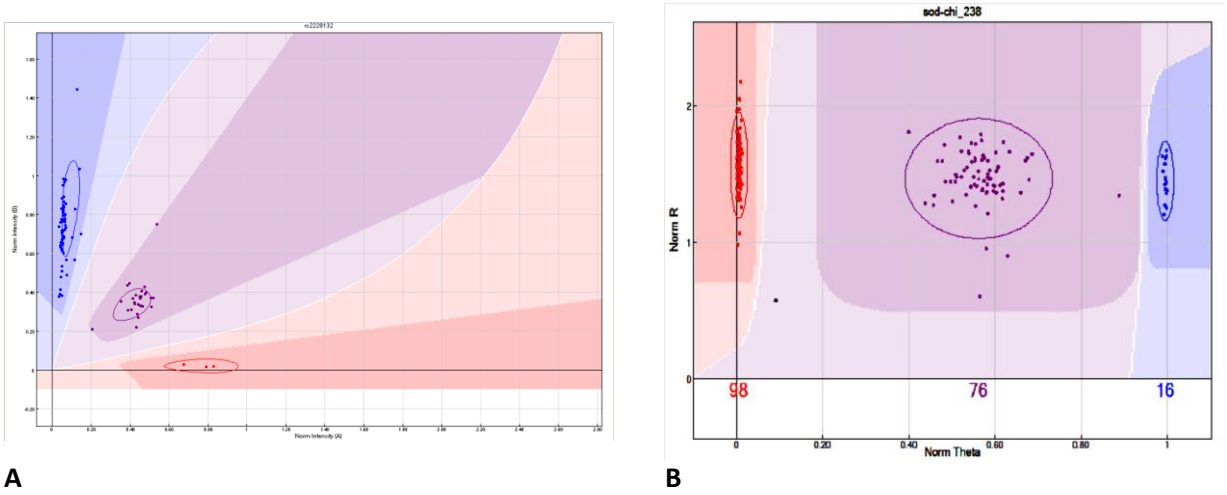


Figure 21. A) Samples clusterization according to their genotypes: intensity values are not normalized. **B)** Normalized graphical representation: the theta angle between the sample and the x axis is converted into a value in the x axis and on the y axis is reported the fluorescence intensity as a theta-normalized angle function.

According to the fluorescence distribution three clusters are shown: usually, homozygous genotypes creates vertical clusters while heterozygous ones generate a more spread cluster. Ideally, Theta angle have to assume a value of 0 for the AA homozygous genotype (red dots), 0.5 for the AB heterozygous genotypes (purple dots) and 1 for BB homozygous genotypes (blue dots). Generally, an analysis is considered good if homozygous dots lie between 0 and 0.2 and between 0.8 and 1 (this means that BeadXpress Reader detects a 20% fluorescence from one dye and the 80% from the other one), while heterozygous dots are positioned between 0.2 and 0.8 on the x axis. A good Norm R value is around 1: conventionally, if it is lower than 0.4 the analysis is considered failed (Fig. 22).

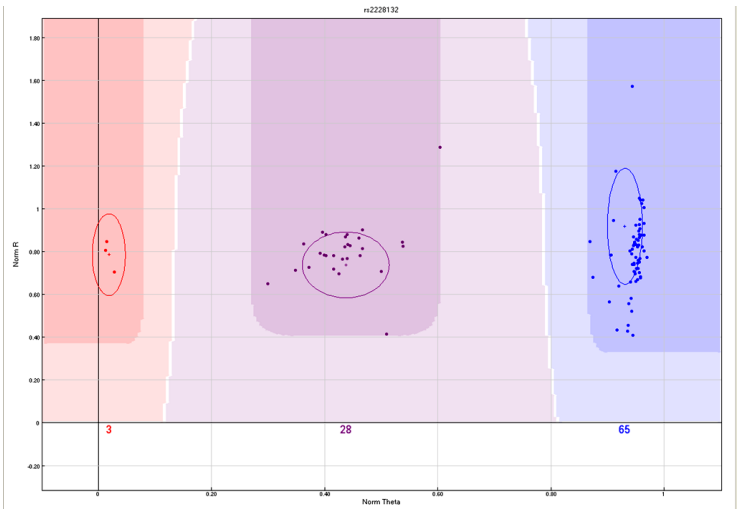


Figure 22. Example of a good result: Norm R is higher than 0.4 and heterozygous dots lie between 0.2 and 0.8 while the homozygous dots have values lower than 0.2 and higher than 0.8.

Based on the obtained clusterization quality, the software assigns to each SNP a score called "gene train score" ranging from 0 to 1 (index of an optimal clusterization).

Three main areas characterize the graphic interface of the software:

- the GenoPlot allowing a graphical plotting of the obtained results, as previously described.
- the "sample table", which lists the sample ID of every analyzed sample. Each sample is associated to the "Call Rate" value, representing the percentage of sample calling among all the analyzed SNPs. This value ranges from 0 to 1, in which 0 means that for the considered sample the software was not able to call any SNP, while 1 means that the genotype of every SNP have been determined for the sample.
- the "full data table" in which all the analyzed SNPs are listed. By selecting one SNP, the user can visualize in the GenoPlot the related clusterization for all the analyzed samples. In this table, each SNP is associated with a quality score, the "gene train score", which depends on sample clusterization, on the SNP final score obtained in the design phase, and on the fluorescence intensity values.

Data analysis

Genome Studio software performs a basic data analysis, then operators can improve call rates and evaluate assay performance, sample quality, and locus performance by following simple guidelines.

Analysis begins with an overall evaluation of the assay performance and determination of which samples, if any, require reprocessing or removal. Clustering should be done after inclusion of reprocessed samples and removal of failed or suboptimal samples, allowing for a more detailed evaluation of sample quality. Each locus can then be evaluated for editing or zeroing (excluding) to optimize call rates. In particular, these parameters need to be analyzed:

1. *controls*

   48 sample-dependent, sample-independent, and contamination controls are all built into the GoldenGate assay. These controls provide a way to assess the overall performance of samples, reagents, equipment, and BeadChips. During preliminary sample quality evaluation, samples falling outside the expected performance parameters should be highlighted for additional analysis. Failure in these controls could indicate a processing failure in a specific step or poor DNA quality.

2. *gencall score*

   Before evaluating the quality of SNP clusters, it is important to highlight samples that have poor performance in the genotyping assay. The GenCall score is a quality metric, ranging from 0–1, calculated for each genotype (data point). GenCall scores generally decrease in value the farer a sample data point is from the center of its cluster.

   Each SNP is evaluated based on the angle, dispersion, and overlap of clusters and intensity.

Problematic samples are identified by a scatter plot of the call rate as a function of the 10% GenCall score (10% GC or p10 GC) (Fig. 23).

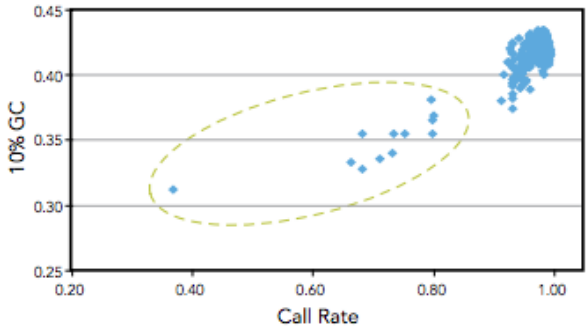


Figure 23. Poorly performing samples are obvious outliers from the majority of samples when 10% GC Score is plotted against sample call rate (green oval).

Poorly performing samples—those with low sample call rates, low 10% GC scores, or outliers from the main population cluster—should be considered for reprocessing or exclusion from the project.

3. *SNP cluster position*

To identify loci that need to be manually edited or zeroed (excluded), metrics listed in the SNP Table are used. These metrics are based on all samples for each locus, providing overall performance information for each locus. To find loci that might need to be edited or removed, data are sorted one column at a time, exploring values at the extremes of the ranges and defining "hard cutoff" and "grey zones". The hard cutoff should be defined as the level, below or above, at which the majority of loci are unsuccessful and should be zeroed. The grey zone should be defined to contain loci that are 80–90% successful and can be improved by manual editing. The upper limit (or lower limit) of the grey zone is the point at which all loci are successful. SNPs falling in the grey zone should be either zeroed or manually edited by moving cluster positions. The meters to be considered are the following:

a. Cluster separation: Cluster Sep measures the separation among the three genotype clusters in the theta dimension and varies from 0–1. Individual SNPs have to be evaluated for overlapping clusters, starting with those having low Cluster Sep. If clusters are well separated, the SNP can be manually edited. SNPs with overlapping clusters should be zeroed.

b. SNP call frequency: Call Freq is the proportion of all samples at each locus with call scores above the no-call threshold. The value varies from 0–1. SNPs have to be evaluated starting with those having low Call Freq values. Zero the SNP if the low call frequency cannot be attributed to a potential biological effect, such as a chromosomal deletion, in a subset of samples.

c. AB Mean for intensity and theta:

   i. AB R Mean is the mean normalized intensity (R) of the heterozygote cluster. This metric helps identify SNPs with low intensity data and has values increasing from 0. SNPs have to be evaluated from low to high AB R Mean and zero any SNPs with intensities too low for genotypes to be called reliably.

   ii. AB T Mean is the mean of the normalized theta values of the heterozygote cluster. This value ranges from 0–1. SNPs with AB T Mean ranging from 0–0.2 and 1–0.8 (or more, if necessary) have to be evaluated to identify SNPs where the heterozygote cluster has shifted toward the homozygotes. If clusters can be reliably separated the SNP could be manually edited; otherwise the locus should be zeroed.

The identification of problematic samples and loci in a systematic manner ensures optimal final data quality from the GoldenGate genotyping assay.

4.3 CLINICAL DATA COLLECTION AND ELABORATION

Patients' clinical data have been collected by oncologists using the suitable created Case Report Form (CRF) (see results).

All personal and clinical data were catalogued in appropriate databases, prepared in accordance with the Privacy Policy, in order to be associated with genetic data.

4.4 RESPONSE AND TOXICITY EVALUATION

Tumor response to hormone therapy treatment was based on the measurement of metastatic sites at enrollment and evaluated every 8 weeks of treatment for at least 24 weeks. Metastatic sites were defined as target measurable lesions and non-target lesions according to RECIST criteria version 1.0 (Therasse *et al.*, 2000). The methods used for assessing the tumor response were: CT scan, radiography, scintigraphy, nuclear magnetic resonance, objective examination, other (i.e. PET).

Tumor response was classified in complete response (CR), partial response (RP), stable disease (SD) and progression disease (PD) according to RECIST criteria version 1.0 (Therasse *et al.*, 2000). Responses were grouped in different manners to evaluate two clinical parameters of response: Response Rate (RR) and Clinical Benefit (CB). For the RR, patients having a CR or a PR were compared with patients showing a SD or PD, while for CB assessment, patients with CR, PR or SD were compared with patients which experienced a PD.

Safety was assessed every 8 weeks for at least 24 weeks, through physical evaluation, laboratory tests, and ECG according to NCI-CTC.

4.5 STATISTICAL ANALYSIS

For each SNP the deviation from Hardy-Weinberg equilibrium was assessed with the http://www.oege.org/software/hwe-mr-calc.shtml program, freely available on line.

For the statistical association between polymorphisms and tumor response the two sided Fisher's exact test was used for obtaining Odds Ratios (OR) and 95% Confidence Intervals (CI). P values ≤ 0.05 were considered statistically significant. The best clinical response obtained during the exemestane treatment was the end-point employed to perform statistical associations.

For this thesis two genetic models were investigated: dominant and recessive. These models require the comparison between patients grouped according to their genotypes, in which the heterozygous group is alternatively merged into an homozygous one. Regarding the dominant model, given that the wild type

(the most frequent genotype) was AA and taken it as the reference category, the compared groups were AA vs AB + BB. In the recessive model, instead, the reference category comprised patients carrying at least one wild type allele (AA + AB), and it was compared to the BB group. For each SNP the most significant genetic model was considered.

To investigate the role of polymorphisms in influencing OS and TTP the Kaplan-Meier product-limit method and the log-rank test statistic were used. The first method was used to trace the curve showing the OS or the TTP and to calculate the median TTP or survival. The log-rank test has been used to compare the curves obtained for each genotype, and to test their statistical difference. The determination of the relative risk of death or progression in patients with different genotype asset was performed by univariate analysis and the results are reported as Hazard Ratios (HR) and the related 95% CI. The OS was evaluated as the months elapsed from the starting date of therapy until the date of death or last control, and the TTP was calculated as the months elapsed from the date of the first exemestane administration until the date of the determination of recurrence or last control.

The software used for statistical analysis were GraphPad InStat® version 3.10 (GraphPad Software, San Diego, CA, USA), STATISTICA™ version 7.1 (StatSoft, Tulsa, OK) and R (http://www.R-project.org).

5.RESULTS

5.1 PATIENTS' CHARACTERISTICS

For this multicenter study, sponsored by the CRO-National Cancer Institute of Aviano, metastatic or locally advanced BC patients treated with exemestane as a first line hormone therapy were enrolled from 2007 to 2012.

Patients were prospectively accrued from several Italian centers and provided informed consent to genetic and kinetic analyses by signing the forms approved by the local Ethical Committee.

18 of the enrolled patients were excluded from the analyses due to the lack of the biological material requested for the study.

Patients provided two aliquots of whole blood and underwent the pharmacogenetics analyses (Table 7).

All the biological material was stored at a temperature of -80 ° C.

Biological samples and patients eligibility	N°	%
PATIENTS		
N° patients considered for the thesis	275	100%
BIOLOGICAL SAMPLES		
Whole blood	275	100%
ANALYSES		
DNA extraction	275	100%
CYP19A1 Ex11 +410A/C genotyping	275	100%
SNPs under the protocol genotyping	>273	>99.27%
Additional SNPs genotyping	>273	>99.27%

Table 7. Patients' eligibility, biological samples provided, genetic analyses performed.

For this PhD thesis a preliminary subpopulation of 275 patients was considered. Statistical analyses in order to associate genetic and clinical data have been conducted on this selected cohort. The median age of patients was 71 years (35 to 93 years) and the median follow up was 35 months (2 to 153 months) but the complete baseline patients characteristics are listed in Table 8.

Characteristic	No.	%
TOTAL	275	
AGE, YEARS		
Median (range)	70 (35-93)	
FOLLOW UP (MONTHS)		
Median (range)	35 (2-153)	
SEX		
Female	274	99.6%
Male	1	0.4%
STAGE AT DIAGNOSIS		
I	35	12.7%
II	85	30.9%
III	68	24.8%
IV	85	30.9%
Unknown	2	0.7%
SURGERY		
Yes	217	78.9%
Radical	123	44.7%
Partial	84	30.6%
Both	10	3.6%
No	58	21.1%
DOMINANT METASTATIC SITE		
Visceral (no epatic lesions)	125	45.5%
Visceral (+ 1-3 epatic lesions)	41	14.9%
Visceral (+ 4-5 epatic lesions)	1	0.4%
Bone only	85	30.9%
Bone + soft tissue	9	3.3%
Soft tissue	14	5.0%
NUMBER OF METASTATIC SITE AT RECRUITMENT		
1	78	28.4%
2	74	26.9%
3	71	25.8%
4	31	11.3%
5	21	7.6%
ER/PgR STATUS		
ER+/PgR+	195	70.9%
ER+/PgR-	74	26.9%
ER-/PgR+	5	1.8%
ER-/PgR-	1	0.4%

Characteristic	No.	%
PRIOR TREATMENTS		
Prior radiotherapy	104	37.8%
Previous systemic therapy	189	68.7%
Chemotherapy only	42	15.3%
Hormone therapy only	43	15.6%
Both chemotherapy and hormone therapy	104	37.8%
Previous chemotherapy	140	50.9%
Adjuvant treatment only	114	41.5%
Metastatic disease only	17	6.2%
Both adjuvant and metastatic	9	3.3%
Previous hormone therapy	146	53.1%
Previous adjuvant tamoxifen only	87	31.6%
Previous adjuvant AI only	29	10.6%
Both previous adjuvant AI and tamoxifen	30	10.9%

Table 8. Baseline characteristics of eligible patients.

5.2 DEVELOPMENT OF PHARMACOGENETIC METHODS FOR POLYMORPHISMS GENOTYPING

Four methodologies have been set up and employed for the genotyping analyses. Three of them (Pyrosequencing, TaqMan® Assay and Fragment Analysis) were applied for the investigation of the candidate SNPs object of this study, while the GoldenGate Assay was mainly used to validate the results obtained with the above mentioned techniques and to implement the number of investigational variants. For each polymorphism the most appropriate technique, based on the best result obtained in the setting up process, was chosen.

1. Polymorphisms analyzed by Pyrosequencing

Pyrosequencing has been employed for the analyses of the following 3 polymorphisms:

- CYP19A1_47T/C (rs700519),
- CYP3A4*1B_-392A/G (rs2740574),
- RIZ1_delP704 (rs2308040).

All the primers necessary for these analyses were designed by using the "PSQ Assay Design" software. This technique required, for each sample, the preparation of a PCR reaction mixture of 49 µL which was added with 1 µL of genomic DNA. The PCR amplification was verified by the electrophoretic run in a 3% agarose gel before the Pyrosequencing analysis. All the sequence primers employed had a concentration of 2µM.

For each polymorphism the genic portion containing the polymorphic nucleotides was identified, in particular:

- ✓ **CYP19A1_47T/C (rs700519):** the sequence of interest is reported below. In yellow are highlighted the PCR primers' bound sequences, in green the genic portion recognized by the sequence primer, and in red the polymorphic nucleotides:

  TTTTTTTCCAGCAAGGATTTGAAAGATGCCATAGAAGTTCTGATAGCAGAAAAAAGA**[C/T]**GCAGGATTTCCACAG AAGAGAAACTGGAAGAATGTATGGACTTTGCCACTGAGTTGATTTTAGCAGAGGTACTGACCTGAACTAACTGTAAT TC

  In Table 9 are reported the primers sequences, the reaction mix and the PCR conditions which gave the best result in the setting up procedure.

SNP	PRIMER SEQUENCES PCR PRODUCT LENGHT	REACTION MIXTURE	PCR CONDITIONS	
CYP19A1_47T/C (rs700519)	FRW:	1X reaction buffer		
	Biot-TTGAAAGATGCCATAGAAGT	2.0 mM $MgCl_2$	95°C for 10 min	
	REV:	0.125 mM dNTPs	Td = 95°C for 30 sec	38 cycles
	AACTCAGTGGCAAAGTCC	0.2 µM primer FRW/REV	Ta = 57.5°C for 30 sec	
	SEQ REV:	1U Taq Flexi	Te = 72°C for 30 sec	
	CTTCTGTGGAAATCCTG	Reaction Volume:	72°C for 10 min	
	PCR product length: 95 bp	49µL MIX + 1µL DNA		

Table 9. FRW = forward primer, REV = reverse primer, SEQ REV = reverse sequence primer, Biot = biotin, Td = denaturation temperature; Ta = annealing temperature, Te = elongation temperature.

This was a reverse assay, so the pyrograms refers to the complementary sequence respect to the above reported one (Fig. 24-25).

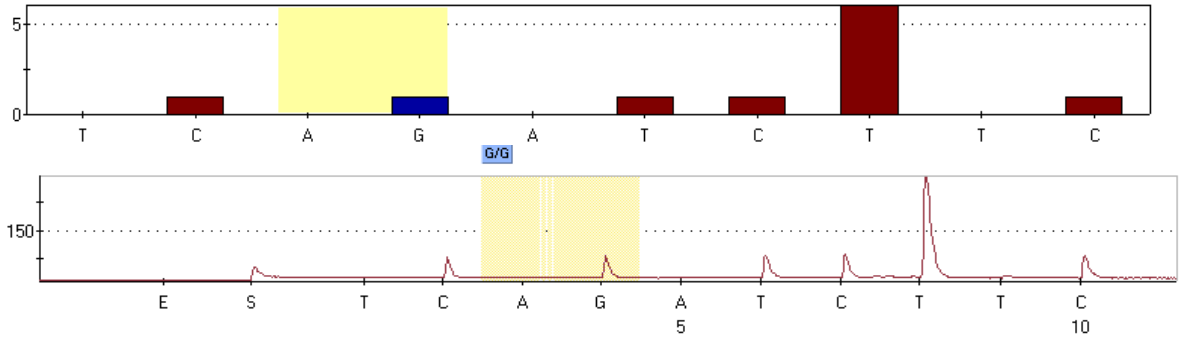


Figure 24. Histogram and pyrogram of the GG genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

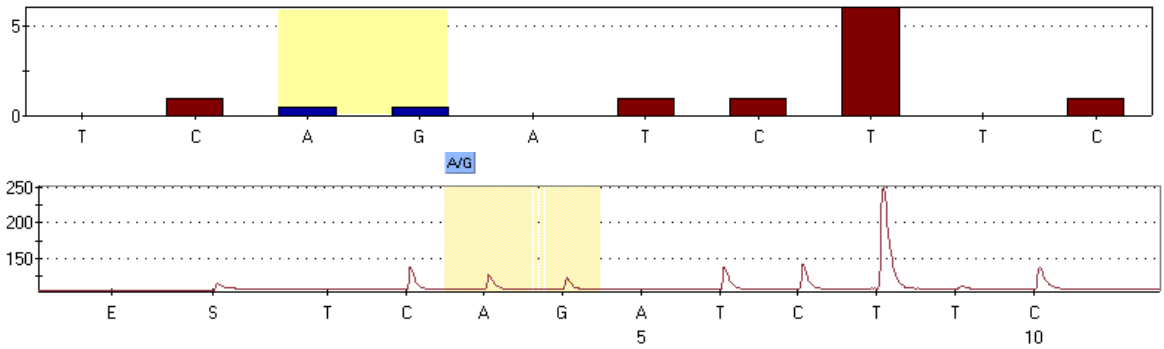


Figure 25. Histogram and pyrogram of the AG genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

✓ **CYP3A4*1B_-392A/G (rs2740574):** the sequence of interest is reported below. In yellow are highlighted the PCR primers bound sequences, in green the genic portion recognized by the sequence primer, and in red the polymorphic nucleotides:

GCACACTCCAGGCATAGGTAAAGATCTGTAGGTGTGGCTTGTTGGGATGAATTTCAAGTATTTTGGAATGAGGACA GCCATAGAGACAAGGGCA[A/G]GAGAGAGGCGATTTAATAGATTTTATGCCAATGGCTCCACTTGAGTTTCTGATA AGAACCCAGAACCCTTGGACTCCCCAGTAACATTGATTGAGTTGTTTATGATACCTCATAGAATATGAACTCAAAGG AGG

In Table 10 are reported the primers sequences, the reaction mix and the PCR conditions which gave the best result in the setting up procedure.

SNP	PRIMER SEQUENCES PCR PRODUCT LENGHT	REACTION MIXTURE	PCR CONDITIONS
CYP3A4*1B_-392A/G (rs2740574)	FRW: ATCTGTAGGTGTGGCTTGT	1X reaction buffer 2.5 mM $MgCl_2$	95°C for 10 min
	REV: Biot-GGGTTCTGGGTTCTTATCA	0.125 mM dNTPs 0.2 µM primer FRW/REV	Td = 95°C for 30 sec Ta = 57°C for 30 sec } 38 cycles
	SEQ FRW: CCATAGAGACAAGGGCA	1U Taq Gold Reaction Volume:	Te = 72°C for 30 sec 72°C for 7 min
	PCR product length: 140bp	49µL MIX + 1µL DNA	

Table 10. FRW = forward primer, REV = reverse primer, SEQ FRW= forward sequence primer, Biot = biotin, Td = denaturation temperature; Ta = annealing temperature, Te = elongation temperature.

Here are reported the pyrograms relative to the CYP3A4*1B_-392A/G (rs2740574) assay (Fig.26-27):

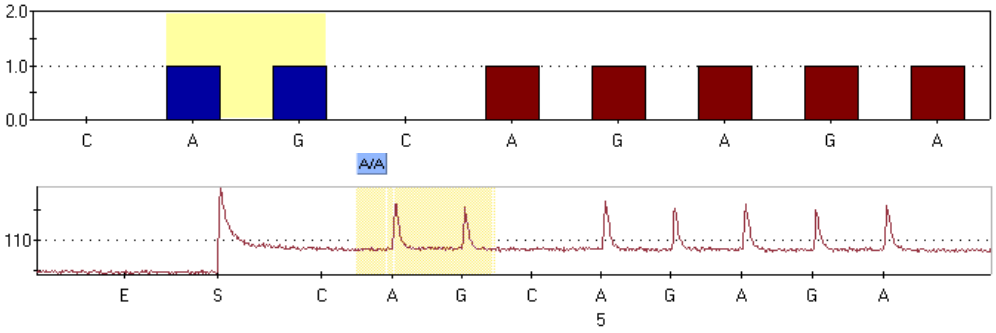


Figure 26. Histogram and pyrogram of the AA genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

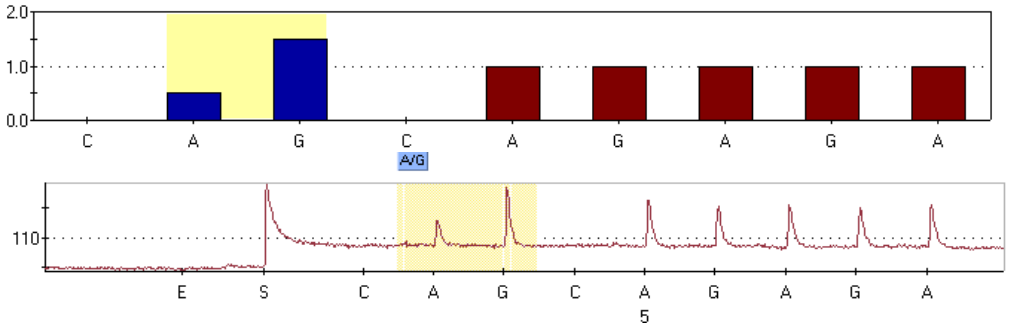


Figure 27. Histogram and pyrogram of the AG genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

- ✓ **RIZ1_delP704 (rs2308040):** this polymorphism consisted on a 3 bp deletion (CTC), corresponding to the omission of the proline residue in the position 704 of the protein. Depending on the alleles, 3 genotype have been considered: "ins/ins" if there is no deletion, "ins/del" if the deletion occurred only in one allele (heterozygous genotype) and "del/del" in the case of homozygous deletion.

  For this polymorphism the sequence of interest is reported below. In yellow are highlighted the PCR primers bound sequences, in green the genic portion recognized by the sequence primer, and in red the polymorphic nucleotides:

  GGGCAGGACTGTGCTGTGGAGAACTGGGAGGAGAGCTGGTCCGCCTCTTAAACCTACTTGAGGTCACAGGCAACAT TGATGCAGGAGCAGACACACAAACAGGACCTAATTTAGCTATTTCAGTTGCTGAAATCCCTGCAGGAGGAG**[-/GAG]**TTA GTTTATCTTGGGTTTGAAGAAGTTGTTTGAGCTTTGATGACAAATAAACACTTTTCTCTTTGTGGAAAGACACTGCCT CTGTTGTTGATATGCTAAGAGGAAGACTTAAAGAGCAAGAGGGGACCATGGGGTCAGAGTC

  In Table 11 are reported the primers sequences, the reaction mix and the PCR conditions which gave the best result in the setting up procedure.

SNP	PRIMER SEQUENCES PCR PRODUCT LENGHT	REACTION MIXTURE	PCR CONDITIONS
RIZ1_delP704 (rs2308040)	FRW: TCACAGGCAACATTGATGCA	1X reaction buffer	
		2.5 mM $MgCl_2$	95°C for 10 min
	REV: GTCCCCTCTTGCTCTTTAAGTC	0.125 mM dNTPs	Td = 95°C for 30 sec
		0.2 µM primer FRW/REV	Ta = 59 °C for 30 sec } 40 cycles
	SEQ REV: TTCAAACCCAAGATAAACTA	1U Taq Gold	Td = 72°C for 30 sec
		Reaction Volume:	72°C for 10 min
	PCR product length: 208 (del) bp 211 (ins) bp	49µL MIX + 1µL DNA	

Table 11. FRW = forward primer, REV = reverse primer, SEQ REV= reverse sequence primer, Td = denaturation temperature; Ta = annealing temperature, Te = elongation temperature.

This was a reverse assay, so the pyrograms refers to the complementary sequence respect to the above reported one (Fig. 28-30):

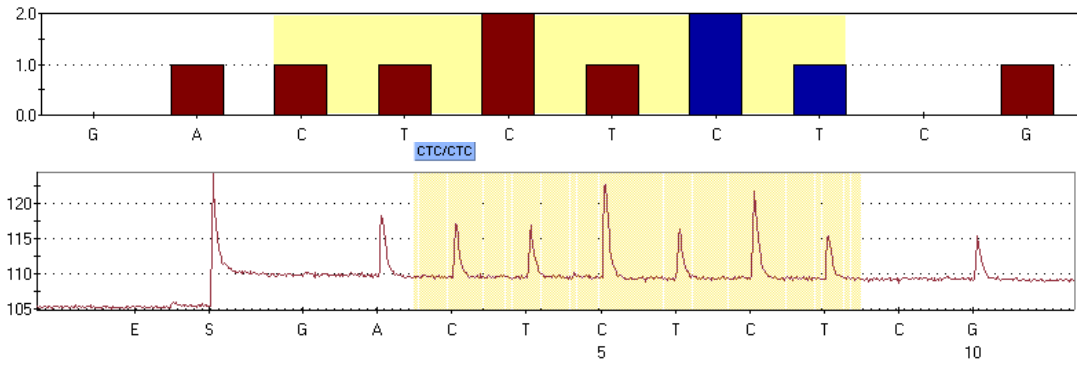


Figure 28. Histogram and pyrogram of the ins/ins genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

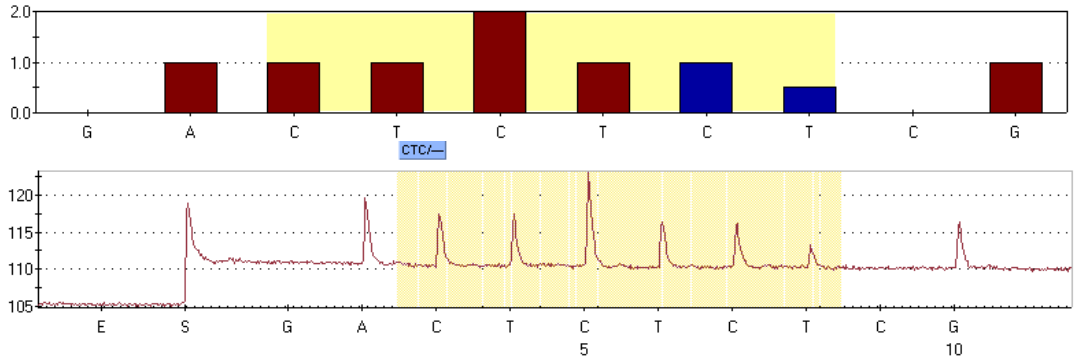


Figure 29. Histogram and pyrogram of the ins/del genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

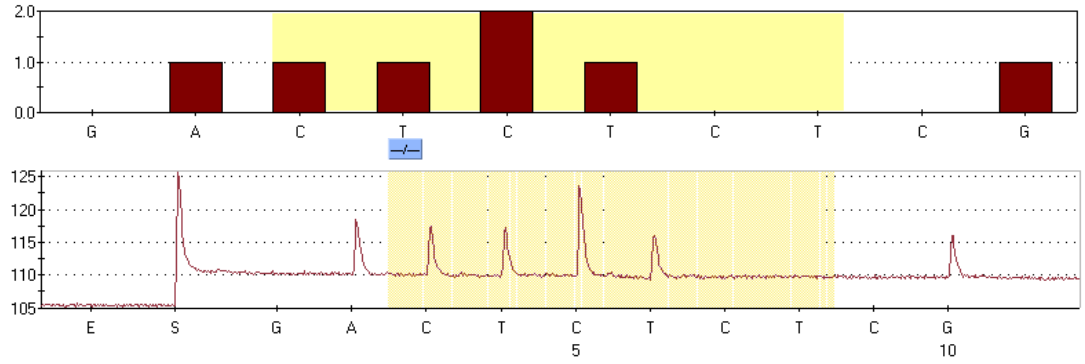


Figure 30. Histogram and pyrogram of the del/del genotype. Histogram represents the expected relative peaks height seen in the pyrogram, according to the number of the nucleotides added. The higher the histogram, the greater the number of inserted nucleotides. In the yellow area are represented the polymorphic bases.

2. Polymorphisms analyzed by TaqMan ® Assay

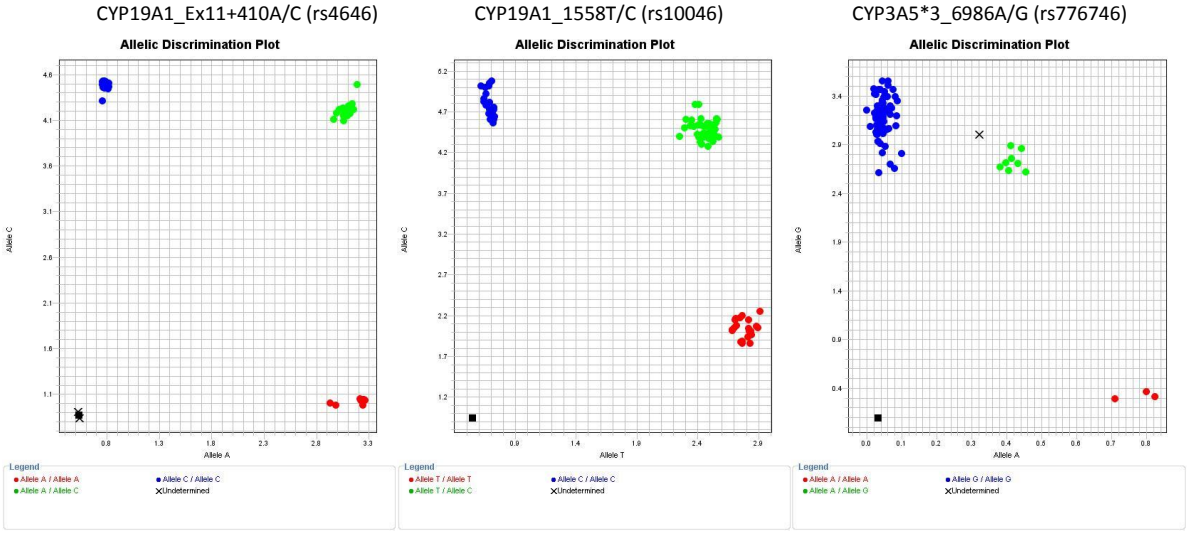
CYP19A1_Ex11+410A/C (rs4646), CYP19A1_1558T/C (rs10046), CYP3A5*3_6986A/G (rs776746), COMT_12A/G (rs4680), ESR1_497T/C (rs2234693), ESR1_256A/G (rs9340799), ESR2_1082A/G (rs1256049), ESR2_1730A/G (rs4986938), CYP17A1_27A/G (rs743572) and CYP1B1*3_4326G/C (rs1056836) were analyzed by pre-designed TaqMan SNP genotyping assays according to the manufacturer's instructions. For each analysis was required a reaction mixture composed by PCR primers (forward and reverse), fluorescent allele-specific probes (SNP assay, Applied Biosystems), and a universal PCR Master Mix containing dNTPs, Taq Polymerase, $MgCl_2$, and properly concentrated salts in a buffered solution. With the high specificity of this technology it is possible to obtain an allelic discrimination in a very low volume of reaction. Samples were analyzed by adding 1 µL of genomic DNA to 5 µL of reaction mixture, according to the best result obtained from the setting up process.

The following Table 12 illustrates the SNPs analyzed by TaqMan ® Assay, the specific code of each assay, the probe concentrations and the reaction volume employed.

SNP	Assay type ID	Probe concentration	Reaction volume
CYP19A1_Ex11+410A/C (rs4646)	C_8234730_1_	20X	5 µL
CYP19A1_1558T/C (rs10046)	C_8234731_30	10X	5 µL
CYP3A5*3_6986A/G (rs776746)	C_26201809_30	20X	5 µL
COMT_12A/G (rs4680)	C_25746809_50	20X	5 µL
ESR1_497T/C (rs2234693)	C_3163590_10	40X	5 µL
ESR1_256A/G (rs9340799)	C_3163591_10	40X	5 µL
ESR2_1082A/G (rs1256049)	C_7573265_1_	40X	5 µL
ESR2_1730A/G (rs4986938)	C_11462726_10	40X	5 µL
CYP17A1_27A/G (rs743572)	C_2852784_30	20X	5 µL
CYP1B1*3_4326G/C (rs1056836)	C_3099976_30	20X	5 µL

Table 12. SNPs analyzed by TaqMan ® Assay.

In Fig. 31 are reported the graphical visualization of the results obtained from the analysis of each SNP.



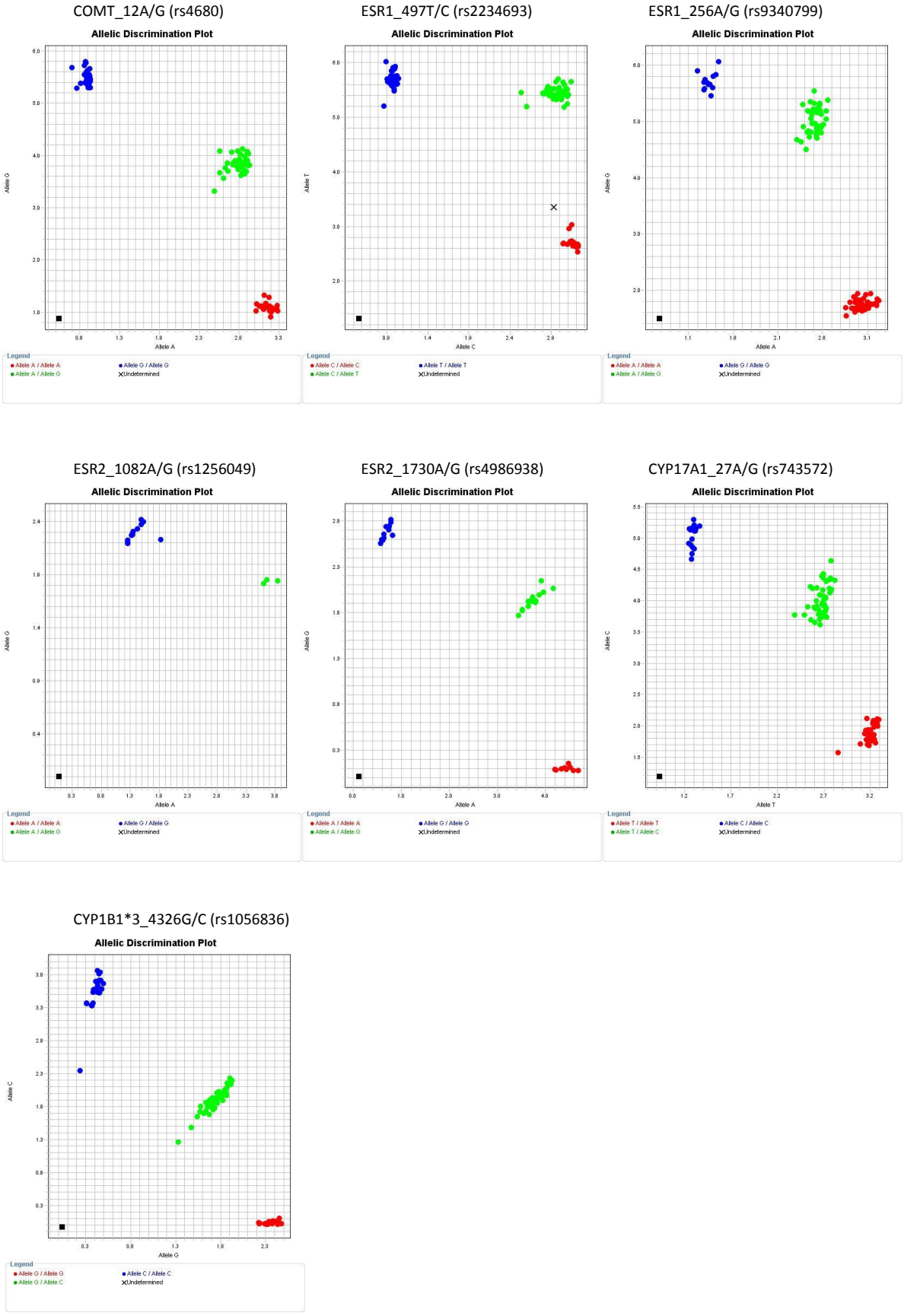


Figure 31. Graphical representation of the results obtained with TaqMan® Assay.

3. Polymorphisms analyzed by Automated Fragment Analysis

CYP19A1_(TTTA)n (rs60271534) and UGT1A1*28_ TA(6/7) (rs8175347) were analyzed by Fragment Analysis. This technique is based on capillary gel electrophoresis and compares the amplified product weight with an internal standard added to each sample.

First of all, the gene fragment containing the repeats was amplified by PCR. To do so, a mixture composed by reaction buffer, $MgCl_2$, dNTPs, forward and reverse primers, Taq polymerase, and water to reach a final volume of 49 µL, was prepared. One of the primers was covalently bounded with HEX. The PCR amplification was verified by an electrophoretic run in a 3% agarose gel before the Fragment Analysis. Based on the intensity of each band on agarose gel, samples were accordingly diluted to reach a final concentration ranging from 1:100 to 1:200 of the original one. 1 µL of each diluted sample was later added to 15 µL mix composed by 14.5 µL of formamide and 0.5 µL of Internal Lane Size Standard (GS 400HD ROX DYE).

- ✓ **CYP19A1_(TTTA)n (rs60271534):** this is a STR polymorphism with a number of TTTA repetitions ranging from 7 to 13, but statistical analysis was done considering long (L) allele (>7 TTTA repeats) and short (S) allele (≤7 TTTA repeats), so generating the genotypes SS, LS and LL.

  For this STR polymorphism the sequence of interest is reported below. In yellow are highlighted the PCR primers bound sequences and in red the polymorphic nucleotides:

  CTGGAAAACAACTCGACCCTTCTTTATGAAAGGTAAGCAGGTACTTAGTTAGCTACAATCTTTTTTGTCTATGAATGT GCCTTTTTTGAAATCATATTTTTAAAATAT**[(TAAA)7/(TAAA)13]**TTGAGACAGGCTCTGACTCTATCACCCAGGC

  Table 13 reports the forward and reverse primers sequences, the optimal PCR conditions and the reaction mix employed for the analysis, according to the best results obtained by the setting up procedure.

STR	PRIMER SEQUENCES	REACTION MIXTURE	PCR CONDITIONS
CYP19A1_(TTTA)n (rs60271534)	FRW: TTGTCTATGAATGTGCCTTTT	1X reaction buffer	
		1.5 mM $MgCl_2$	95°C for 10 min
		0.125 mM dNTPs	Td = 95°C for 1 min
	REV: HEX-CTGGGTGATAGAGTCAGAGC	0.5 µM primer FRW/REV	Ta = 59 °C for 1 min
		1U Taq Gold	Te = 72°C for 1.30 min
		Reaction Volume: 50µL	72°C for 10 min

40 cycles

Table 13. FRW = forward primer, REV = reverse primer, Td = denaturation temperature; Ta = annealing temperature, Te = elongation temperature.

In Table 14 are listed the molecular weights, expressed in base pairs (bp), of each fragment analyzed, while Fig. 32 illustrates the electropherograms of the peacks corresponding to the different fragments lengths.

N° OF TTTA REPEATS	MOLECULAR WEIGHT (BP)
7	95
8	99
9	103
10	107
11	111
12	115
13	119
14	123

Table 14. Number of TTTA repeats and the corresponding molecular weight expressed in bp.

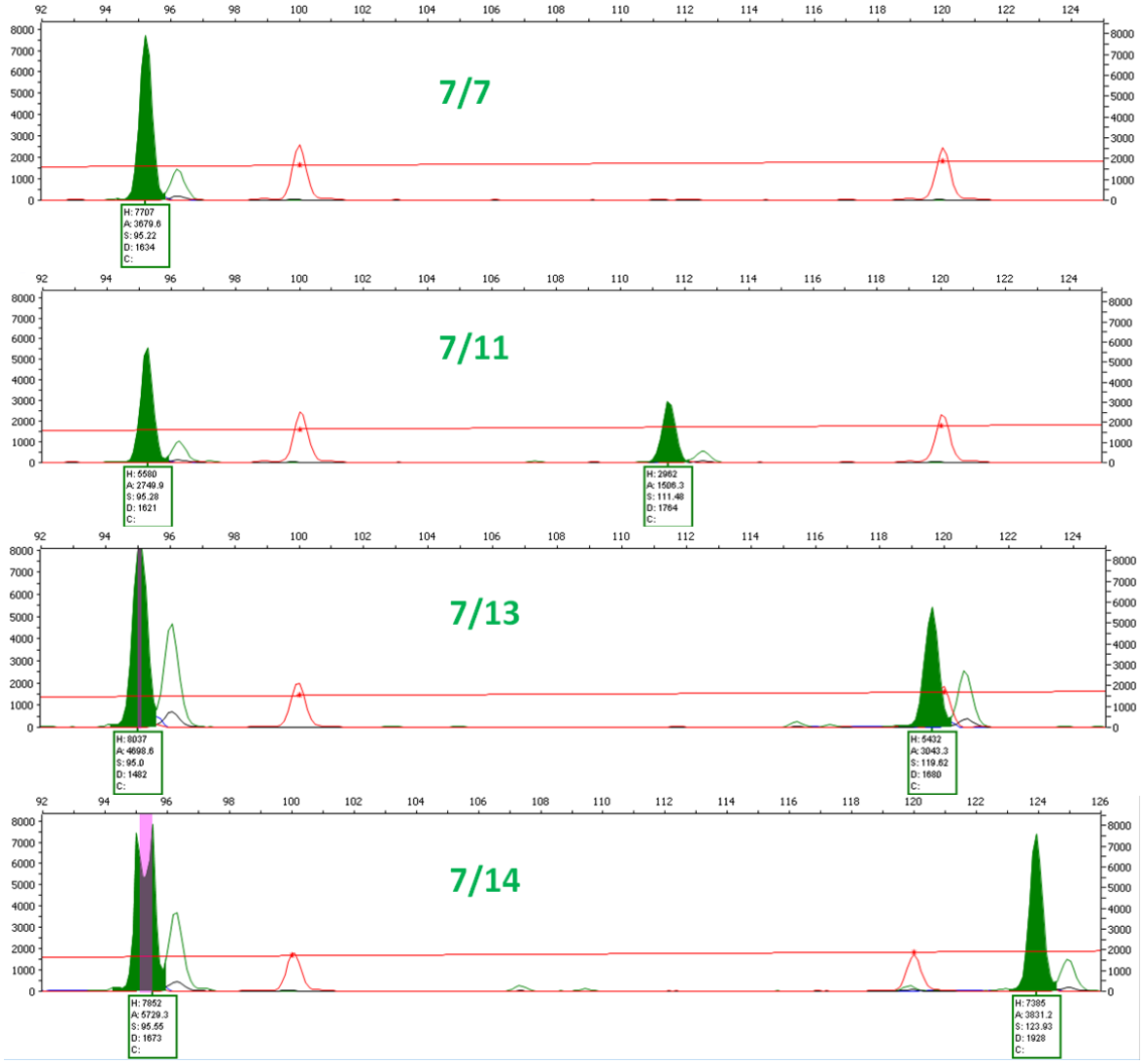


Figure 32. Electropherograms corresponding to the 7/7; 7/11; 7/13 and 7/14 genotypes.

The number of repetitions described in literature ranges from 7 to 13, but in our cohort, we found a patient carrying a 14 (TTTA) repeats allele, as illustrated in the lowest panel of Fig 32.

To make sure that the result observed was not an artifact, the analysis was twice repeated with the Fragment Analysis and was later validated by the Sanger Sequencing.

Briefly, Sanger sequencing (Sanger *et al.*, 1977) is a method of DNA sequencing based on the selective incorporation of chain-terminating modified di-deoxynucleotides (ddNTPs) by DNA polymerase for detection in automated sequencing machines. These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, causing a stop in DNA extension when a modified ddNTP is incorporated. This process generates a pool of DNA fragment with different length, each one terminating with a ddNTP. These are fluorescently labeled with four dye-terminators, each one emitting at different wavelengths. The resulting DNA fragments are denatured both by heat and formammide and subsequently separated performing a capillary electrophoretic run, similarly as for the automated fragment analysis.

To validate the result obtained, a new forward primer was designed by Primer3 software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). The change of the forward primer resulted in a new PCR setup and thus in the change of PCR reaction mixture and conditions ($MgCl_2$ concentration, Ta and duration of each PCR cycle phases as reported in Table 15).

In the following sequence in yellow are highlighted the new PCR primers' bound portion and in red the polymorphic nucleotides:

TCTCGATTCGGCAGCAAACTTGGGCTGCAGTGCATCGGTATGCATGAGAAAGGCATCATATTTAACAACAATCCAGAGCTCTG
GAAAACAACTCGACCCTTCTTTATGAAAGGTAAGCAGGTACTTAGTTAGCTACAATCTTTTTTGTCTATGAATGTGCCTTTTTTG
AAATCATATTTTTAAAATAT**[(TAAA)7/(TAAA)13]**TTGAGACAGGCTCTGACTCTATCACCCAGGCAGGAGTGACCTT

To perform the PCR required for the Sanger Sequencing, a mixture, containing reaction buffer, $MgCl_2$, dNTPs, forward and reverse primers, Taq polymerase, and water to reach a final volume of 49 µL, was prepared.

Table 15 reports the forward and reverse primers sequences, the optimal PCR conditions and the reaction mix employed for the Sanger Sequencing, according to the best results obtained by the setting up procedure.

STR	PRIMER SEQUENCES	REACTION MIXTURE	PCR CONDITIONS
CYP19A1_(TTTA)n (rs60271534)	FRW:	1X reaction buffer	
	TCTCGATTCGGCAGCAAACT	2.5 mM $MgCl_2$	95°C for 10 min
	REV:	0.125 mM dNTPs	Td = 95°C for 30 sec
	CTGGGTGATAGAGTCAGAGC	0.5 µM primer FRW/REV	Ta = 62 °C for 30 sec } 38 cycles

PCR product length: 245 bp	1U Taq Gold	Te = 72°C for 30 sec
	Reaction Volume: 50µL	72°C for 7 min

Table 15. FRW = forward primer, REV = reverse primer, Td = denaturation temperature; Ta = annealing temperature, Te = elongation temperature.

This analysis was performed in multiple independent PCRs and confirmed that the number of repetitions was 14, as illustrated in Fig. 33.

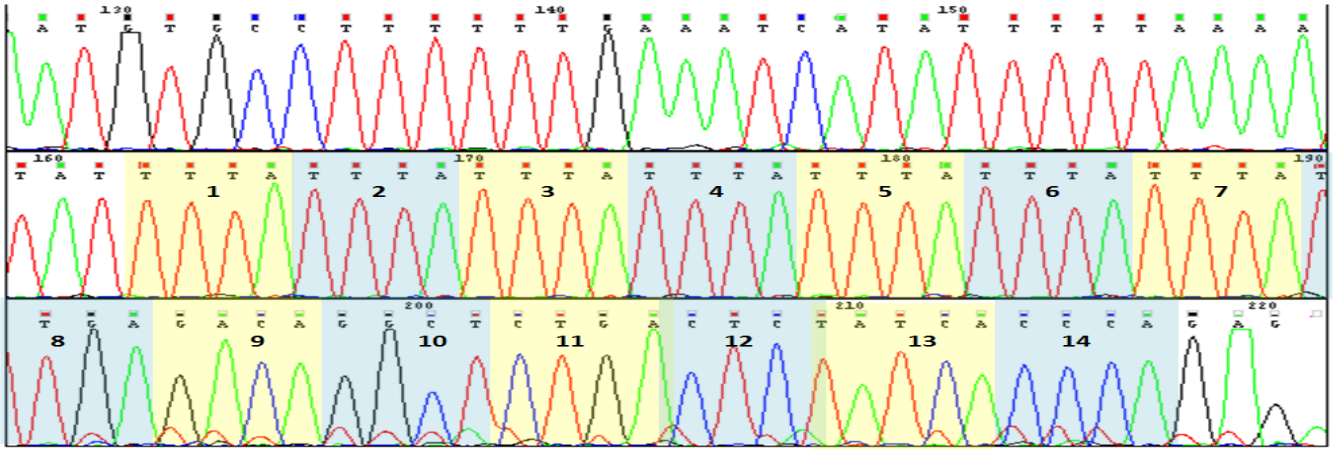


Figure 33. Results obtained with the Sanger Sequencing.

- ✓ **UGT1A1*28_TA(6/7) (rs8175347)**: this is a STR polymorphism characterized by a variable number of dinucleotide (TA) repeats on the promoter region ranging from 5 to 8.

  Even for this polymorphism the statistical analysis was done considering long (L) allele (≥7 TA repeats) and short (S) allele (<7 TA repeats), so generating the genotypes SS, LS and LL.

  For this STR the sequence of interest is reported below. In yellow are highlighted the PCR primers bound sequences and in red the polymorphic nucleotides:

  TGCTACCTTTGTGGACTGACAGCTTTTTATAGTCACGTGACACAGTCAAACATTAACTTGGTGTATCGATTGGTTTTTG
  CCA**[(TA)5/(TA)8]**AGTAGGAGAGGGCGAACCTCTGGCAGGAGCAAAGGCGCCATGGCTGTGGAGTCCCAGG

  Table 16 reports the forward and reverse primers sequences, the optimal PCR conditions and the reaction mix employed for the analysis, according to the best results obtained by the setting up procedure.

STR	PRIMER SEQUENCES	REACTION MIXTURE	PCR CONDITIONS	
UGT1A1*28_TA (rs8175347)	FRW: GTCACGTGACACAGTCAAACATTAAC TTGG	1X reaction buffer		
		3 mM $MgCl_2$	95°C for 10 min	
		0.25 mM dNTPs	Td = 95°C for 30 sec	38 cycles
	REV: HEX-TTTGCTCCTGCCAGAGGTT	0.5 µM primer FRW/REV	Ta = 65.5 °C for 30 sec	
		1U PolyTaq	Te = 72°C for 30 sec	
		Reaction Volume: 20µL	72°C for 10 min	

Table 16. FRW = forward primer, REV = reverse primer, Td = denaturation temperature; Ta = annealing temperature, Te = elongation temperature.

In Table 17 are listed the molecular weights, expressed in base pairs (bp), of each fragment analyzed.

N° OF TA REPEATS	MOLECULAR WEIGHT (BP)
5	94
6	96
7	98
8	100

Table 17. Number of TA repeats and the corresponding molecular weight expressed in bp.

Figure 34 represents the typical UGT1A1 fragment analysis electropherograms and illustrates the difference in the length depending on the number of TA repeats.

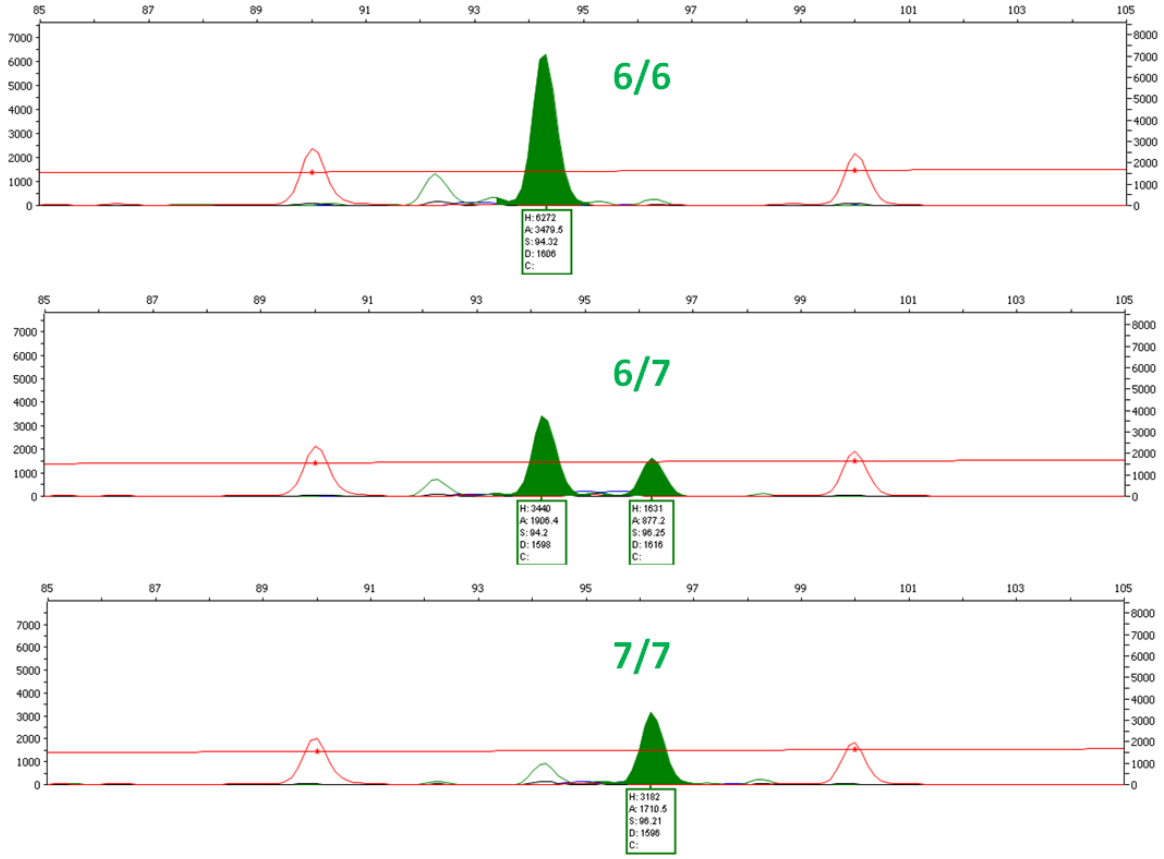


Figure 34. Electropherograms corresponding to the UT1A1 6/6; 6/7; 7/7 genotypes.

4. Polymorphisms analyzed by GoldenGate Assay

Due to the technological implementation occurred in our laboratory with the acquisition of the BeadXpress Reader and with the possibility of using the Illumina GoldenGate assay, it was decided to move, from the analysis of the 15 polymorphisms to a higher through-put investigation with a wider panel of candidate polymorphisms, possibly associated with the clinical outcome of patients.

It has been, thus, obtained a list of 96 SNPs, which was then submitted to Illumina for the synthesis of specific oligonucleotides for the corresponding GoldenGate assays. Two kits for the simultaneous analysis of 96 polymorphisms, which differed by 25 SNPs, were applied. In Table 18 is reported the total list, including the 71 SNPs in common and the two set of 25 polymorphisms peculiar of each kit.

12 out of the 15 polymorphisms originally included in the protocol, (in bold font in the table) were introduced in the panels of analysis as positive controls for the validation of the new method.

Gene	SNP name	rs code	Final Score	Observed allele frequency				1°, 2° or both kits
ABCB1	ABCB1_127A/G	rs2235023	0,58	A	7,1%	G	92,9%	2° kit
	ABCB1_–129T/C	rs3213619	0,52	T	96,2%	C	3,8%	both
	ABCB1_1841T/C	rs12334183	0,66	T	85,3%	C	14,7%	2° kit
	ABCB1_IVS13+24T/C	rs2235033	0,68	T	48,1%	C	51,9%	both
	ABCB1_IVS14+38A/G	rs2235013	0,06	A	51,8%	G	48,2%	both
	ABCB1_IVS4–25T/G	rs2235015	0,05	T	18,6%	G	81,4%	both
	ABCB1_IVS9-44T/C	rs10276036	0,62	T	58,0%	C	42,0%	both
	ABCB1_rs3842A/G	rs3842	0,45	A	86,4%	G	13,6%	both
ACBC1	ABCC1_1062T/C	rs35587	0,60	T	69,5%	C	30,5%	both
	ABCC1_1303T/G	rs60782127	0,60	T	1,3%	G	98,7%	1° kit
	ABCC1_1684T/C	rs35605	0,55	T	18,5%	C	81,5%	both
	ABCC1_2158A/G	rs4148356	0,65	A	0,4%	G	99,6%	1° kit
	ABCC1_–260C/G	rs504348	1,01	C	82,6%	G	17,4%	1° kit
	ABCC1_2965A/G	rs35529209	0,61	A	0,2%	G	99,8%	1° kit
	ABCC1_4002A/G	rs2230671	0,44	A	21,6%	G	78,4%	1° kit
	ABCC1_IVS11-48T/C	rs3765129	0,55	T	13,1%	C	86,9%	both
	ABCC1_IVS18-30C/G	rs2074087	0,50	C	17,0%	G	83,0%	both
	ABCC1_IVS30+18T/C	rs212088	0,49	T	15,5%	C	84,5%	both
	ABCC1_IVS9+8A/G	rs35588	0,60	A	70,7%	G	29,3%	both
ABCC2	ABCC2_1249A/G	rs2273697	1,01	A	22,2%	G	77,8%	both
	ABCC2_-24A/G	rs717620	1,01	A	18,4%	G	81,6%	both
	ABCC2_3662T/A	rs17222723	0,65	T	N.A.	A	N.A.	1° kit
	ABCC2_3972A/G	rs3740066	1,01	A	36,5%	G	63,5%	both
	ABCC2_4544A/G	rs8187710	1,01	A	7,0%	G	93,0%	both
	ABCC2_intron19T/C	rs2002042	1,01	T	21,8%	C	78,2%	both

Gene	SNP name	rs code	Final Score	Observed allele frequency				1°, 2° or both kits
	ABCC2_IVS23+56T/C	rs4148396	0,63	T	37,4%	C	62,6%	both
	ABCC2_IVS26-34T/C	rs17216177	0,68	T	92,3%	C	7,7%	1° kit
ABCG2	ABCG2_1143T/C	rs2622604	1,01	T	22,1%	C	77,9%	both
	ABCG2_16702A/G	rs2046134	0,69	A	2,3%	G	97,7%	both
	ABCG2_34A/G	rs2231137	0,58	A	5,0%	G	95,0%	1° kit
	ABCG2_421A/C	rs2231142	0,59	A	8,1%	C	91,9%	1° kit
	ABCG2_CTCAdel	rs3219191	0,47	I	42,5%	D	57,5%	both
AGXT	AGXT_1142A/G	rs4426527	1,01	A	83,0%	G	17,0%	2° kit
	AGXT_154T/C	rs34116584	0,40	T	15,6%	C	84,4%	2° kit
APE1	APE1_2197T/G	rs1130409	0,68	T	55,1%	G	44,9%	both
ATIC	ATIC_347C/G	rs2372536	1,01	C	65,4%	G	34,6%	both
ATM	ATM_40C/G	rs1800054	0,62	C	98,9%	G	1,1%	1° kit
	ATM_61A/G	rs1801516	1,01	A	13,2%	G	86,8%	both
BAX	BAX_248A/G	rs4645878	0,42	A	N.A.	G	N.A.	both
CCDN1	CCDN1_1722C/G	rs678653	1,01	C	38,2%	G	61,8%	2° kit
	CCDN1_-870A/G	rs9344	0,62	A	50,3%	G	49,7%	2° kit
COMT	**COMT_12A/G**	**rs4680**	**1,01**	**A**	**47,9%**	**G**	**52,1%**	**both**
CYP17A1	**CYP17A1_27A/G**	**rs743572**	**1,01**	**A**	**58,3%**	**G**	**41,7%**	**both**
CYP19A1	**CYP19A1_1558T/C**	**rs10046**	**1,01**	**T**	**51,5%**	**C**	**48,5%**	**both**
	CYP19A1_47T/C	**rs700519**	**0,66**	**T**	**2,8%**	**C**	**97,2%**	**both**
	CYP19A1_ex11+410A/C	**rs4646**	**1,01**	**A**	**28,8%**	**C**	**71,2%**	**both**
CYP1B1	**CYP1B1*3_4326G/C**	**rs1056836**	**0,58**	**G**	**44,2%**	**C**	**55,8%**	**both**
CYP2C8	CYP2C8*1C_370A/C	rs17110453	0,56	A	87,9%	C	12,1%	1° kit
	CYP2C8*4_792C/G	rs1058930	1,01	C	96,5%	G	3,5%	1° kit
CYP2C9	CYP2C9*2_430T/C	rs1799853	1,01	T	10,9%	C	89,1%	1° kit
	CYP2C9*3_1075A/C	rs1057910	1,01	A	N.A.	C	N.A.	1° kit
CYP3A4	**CYP3A4*1B_-392A/G**	**rs2740574**	**1,01**	**A**	**98,0%**	**G**	**2,0%**	**both**
CYP3A5	**CYP3A5*3_6986A/G**	**rs776746**	**1,01**	**A**	**5,1%**	**G**	**94,9%**	**both**
DYPD	DYPD_IVS14+1A/G	rs3918290	0,06	A	0,1%	G	99,9%	both
EDN1	EDN1_61T/G	rs5370	1,01	T	23,7%	G	76,3%	2° kit
EGF	EGF_61A/G	rs4444903	1,01	A	61,7%	G	38,3%	2° kit
EGFR	EGFR_497A/G	rs2227983	0,28	A	25,3%	G	74,7%	2° kit
ERCC1	ERCC1_19007T/C	rs11615	1,01	T	59,0%	C	41,0%	both
	ERCC1_8092T/G	rs3212986	1,01	T	28,3%	G	71,7%	both
ESR1	**ESR1_497T/C**	**rs2234693**	**1,01**	**T**	**55,6%**	**C**	**44,4%**	**both**
ESR2	**ESR2_1082A/G**	**rs1256049**	**1,01**	**A**	**2,4%**	**G**	**97,6%**	**both**
	ESR2_1730A/G	**rs4986938**	**1,01**	**A**	**40,5%**	**G**	**59,5%**	**both**
FGFR4	FGFR4_1217T/C	rs351855	0,42	T	28,2%	C	71,8%	both
FOLR1	FOLR1_1314A/G	rs2071010	0,67	A	7,0%	G	93,0%	both
	FOLR1_1928T/C	rs9282688	0,59	T	2,8%	C	97,2%	both

Gene	SNP name	rs code	Final Score	Observed allele frequency				1°, 2° or both kits
FPGS	FPGS_1994A/G	rs10106	1,01	A	61,7%	G	38,3%	both
GGH	GGH_354T/G	rs719235	1,01	T	27,6%	G	72,4%	2° kit
	GGH_452T/C	rs11545078	0,68	T	10,8%	C	89,2%	both
GSTA1	GSTA1*B_69A/G	rs3957357	0,06	A	N.A.	G	N.A.	1° kit
GSTM3	GSTM3_del/AGG	rs1799735	0,49	I	84,3%	D	15,7%	2° kit
GSTP1	GSTP1_313A/G	rs1695	1,01	A	68,2%	G	31,8%	both
	GSTP1_341T/C	rs1138272	0,63	T	4,5%	C	95,5%	both
hEXO1	hEXO1_Ex12+49T/C	rs4149963	1,01	T	8,4%	C	91,6%	2° kit
hMSH2	hMSH2_IVS12-6T/C	rs2303428	1,01	T	90,8%	C	9,2%	1° kit
hMSH6	hMSH6_556T/G	rs3136228	1,01	T	60,9%	G	39,1%	both
HNF1A	HNF1A_79T/G	rs1169288	0,49	T	63,7%	G	36,3%	2° kit
hOGG1	hOGG1_1245C/G	rs1052133	0,56	C	78,9%	G	21,1%	both
IL6	IL6_-174G/C	rs1800795	1,01	G	68,1%	C	31,9%	both
ITPA	ITPA_94A/C	rs1127354	0,68	A	7,3%	C	92,7%	both
MDM4	MDM4_34091A/C	rs4245739	0,55	A	68,5%	C	31,5%	both
MGMT	MGMT_Ex5-25T/C	rs12917	1,01	T	12,8%	C	87,2%	2° kit
MTHFD1	MTHFD1_1958T/C	rs2236225	0,55	T	43,0%	C	57,0%	both
MTHFR	MTHFR_1298A/C	rs1801131	0,68	A	68,4%	C	31,6%	both
	MTHFR_677T/C	rs1801133	1,01	T	44,1%	C	55,9%	1° kit
MTR	MTR_2756A/G	rs1805087	1,01	A	81,2%	G	18,8%	both
MTRR	MTRR_66A/G	rs1801394	0,63	A	49,6%	G	50,4%	both
NCF4	NCF4_212A/G	rs1883112	1,01	A	N.A.	G	N.A.	2° kit
NOS2A	NOS2A_IVS16+88T/G	rs9282801	1,01	T	39,8%	G	60,2%	2° kit
NOS3	NOS3_786T/C	rs2070744	0,05	T	56,3%	C	43,7%	2° kit
p21	p21_70T/C	rs1059234	0,53	T	N.A.	C	N.A.	1° kit
	p21_98A/C	rs1801270	0,50	A	6,7%	C	93,3%	both
PCFT	PCFT_114A/G	rs17719944	0,59	A	90,7%	G	9,3%	2° kit
RAC2	RAC2_7508T/A	rs13058338	0,58	T	77,7%	A	22,3%	2° kit
RAD51	RAD51_135C/G	rs1801320	1,01	C	7,0%	G	93,0%	1° kit
RFC	RFC_80A/G	rs1051266	1,01	A	46,4%	G	53,6%	1° kit
RIZ1	**RIZ1_delP704**	**rs2308040**	**0,54**	**I**	**38,5%**	**D**	**61,5%**	**both**
SHMT	SHMT_1420T/C	rs2273029	0,51	T	26,1%	C	73,9%	both
SLCO1B1	SLCO1B1_521T/C	rs4149056	0,57	T	84,4%	C	15,6%	both
TP53	TP53_ex4+119G/C	rs1042522	0,51	G	74,1%	C	25,9%	both
	TP53_IVS2+38G/C	rs1642785	0,34	G	N.A.	C	N.A.	1° kit
	TP53_PIN3_IVS3+16bp	rs17878362	0,33	I	N.A.	D	N.A.	1° kit
TP73	TP73_4T/C	rs2273953	0,53	T	21,2%	C	78,8%	1° kit
TYMS	TYMS_1053T/C	rs699517	1,01	T	35,6%	C	64,4%	both
	TYMS_1122A/G	rs2790	1,01	A	77,6%	G	22,4%	both
	TYMS_1494del6	rs16430	0,32	I	64,5%	D	35,5%	both

Gene	SNP name	rs code	Final Score	Observed allele frequency				1°, 2° or both kits
	TYMS_68A/G	rs1059394	1,01	A	35,3%	G	64,7%	1° kit
UGT1A1	UGT1A1*60_-3279A/C	rs4124874	1,01	A	52,7%	C	47,3%	2° kit
	UGT1A1*93_-3156A/G	rs10929302	1,01	A	33,8%	G	66,2%	2° kit
UGT1A7	UGT1A7*4_622T/C	rs11692021	1,01	T	N.A.	C	N.A.	2° kit
UGT1A9	UGT1A9_-440T/C	rs2741045	1,01	T	32,7%	C	67,3%	2° kit
VEGF	VEGF_634C/G	rs2010963	0,07	C	38,9%	G	61,1%	1° kit
XPD	XPD_32C/G	rs238417	0,05	C	42,7%	G	57,3%	2° kit
	XPD_35931T/G	rs13181	1,01	T	58,2%	G	41,8%	both
XPG	XPG_3508G/C	rs17655	1,01	G	19,9%	C	80,1%	both
XRCC1	XRCC1_1449delGGCC	rs3213239	0,33	I	57,0%	D	43,0%	2° kit
	XRCC1_26304T/C	rs1799782	0,06	T	6,1%	C	93,9%	both
	XRCC1_28152A/G	rs25487	1,01	A	33,3%	G	66,7%	both
	XRCC1_Ex9+16A/G	rs25489	0,35	A	6,6%	G	93,4%	both
XRCC3	XRCC3_17893A/G	rs1799796	1,01	A	71,9%	G	28,1%	both
	XRCC3_18067T/C	rs861539	0,41	T	40,0%	C	60,0%	both
	XRCC3_4541A/G	rs1799794	1,01	A	79,1%	G	20,9%	both

Table 18. List of 121 polymorphisms analyzed with the GoldenGate Assay.

5.3 GENOTYPING RESULTS

The genotyping of the original 15 polymorphisms has been completed for all the 275 patients.

Table 19 shows the data of the CYP19A1_Ex11+410A/C (rs4646) SNP, which was the primary genetic end-point of the clinical study. The genotypes distribution of this polymorphism are consistent with data from NCBI (http://www.ncbi.nlm.nih.gov/snp/).

Polymorphism	Total	N°	%	Observed allelic frequency	NCBI allelic frequency
CYP19A1_Ex11+410A/C (rs4646)	**275**				
AA		23	8.4%		
AC		108	39.2%	A = 28.0%	A = 23.5%
CC		144	52.4%	C = 72.0%	C = 76.5%

Table 19. CYP19A1_Ex11+410A/C genotyping data and the comparison between the frequencies observed in the study population and the NCBI ones.

Table 20 refers to the polymorphisms of the secondary end-point of the study.

For each polymorphism, the table shows the total number of patients analyzed, the frequency of the various genotypes and the allele frequencies of both the wild type and the variant allele.

Polymorphism	Total	N°	%	Observed allelic frequency	NCBI allelic frequency
CYP1B1*3_4326G/C (rs1056836)	**275**				
CC		89	32.4%		
GC		127	46.2%	C = 55.5%	C = 55.3%
GG		59	21.4%	G = 44.5%	G = 44.7%
CYP17A1_27A/G (rs743572)	**275**				
AA		95	34.5%		
AG		133	48.4%	A = 58.7%	A = 61.9%
GG		97	35.3%	G = 41.3%	G = 38.1%
CYP19A1_ (TTTA)n (rs60271534)*	**275**				
LL		76	27.6%		
SL		122	44.4%	L = 49.8%	L = NA
SS		77	28.0%	S = 50.2%	S = NA
CYP19A1_47T/C (rs700519)	**275**				
CC		258	93.8%		
TC		17	6.2%	C = 96.9%	C = 96.9%
TT		0	0%	T = 3.1%	T = 3.1%

Polymorphism	Total	N°	%	Observed allelic frequency	NCBI allelic frequency
CYP19A1_1558T/C (rs10046)	**275**				
TT		85	30.9%		
TC		122	44.4%	T = 53.1%	T = 56.6%
CC		68	24.7%	C = 46.9%	C = 43.4%
CYP3A4*1B_-392A/G (rs2740574)	**275**				
AA		265	96.4%		
AG		10	3.6%	A = 98.2%	A = 97.5%
GG		0	0%	G = 1.8%	G = 2.5%
CYP3A5*3_6986A/G (rs776746)	**275**				
GG		248	90.2%		
AG		26	9.4%	G = 94.9%	G = 96.4%
AA		1	0.4%	A = 5.1%	A = 3.6%
COMT_12A/G (rs4680)	**275**				
GG		74	26.9%		
AG		133	48.4%	G = 51.1%	G = 52.2%
AA		68	24.7%	A = 48.9%	A = 47.8%
ESR1_256A/G (rs9340799)	**275**				
AA		109	39.6%		
AG		130	47.3%	A = 63.3%	A = 69.4%
GG		36	13.1%	G = 36.7%	G = 30.6%
ESR1_497T/C (rs2234693)	**275**				
TT		84	30.5%		
TC		138	50.2%	T = 55.6%	T = 59.3%
CC		53	19.3%	C = 44.4%	C = 40.7%
ESR2_1730A/G (rs4986938)	**275**				
GG		91	33.1%		
AG		137	49.8%	G = 58.0%	G = 61.9%
AA		47	17.1%	A = 42.0%	A = 38.1%
ESR2_1082A/G (rs1256049)	**275**				
GG		263	95.6%		
AG		12	4.4%	G = 97.8%	G = 96.9%
AA		0	0%	A = 2.2%	A = 3.1%
RIZ1_delP704 (rs2308040)	**275**				
DD		110	40.0%		

Polymorphism	Total	N°	%	Observed allelic frequency	NCBI allelic frequency
ID		127	46.2%	D = 63.1%	D = 55.0%
II		38	13.8%	I = 36.9%	I = 45.0%
UGT1A1*28_TA(6/7) (rs8175347)**	**273**				
SS		109	39.9%		
SL		130	47.6%	S = 63.7%	S = NA
LL		34	12.5%	L = 36.3%	L = NA

Table 20. Polymorphisms genotyped in this study. S = "short", L = "long" alleles, D = deletion, I = insertion. * S = < 7 TTTA repeats, L = ≥ 7 TTTA repeats. ** S = < 7 TA repeats, L = ≥ 7 TA repeats.

Some of the polymorphisms analyzed were very rare in the population studied.

In particular, it was found no patient with homozygous genotype for the variant allele of the polymorphisms CYP19A1_47T/C (rs700519), ESR2_1082A/G (rs1256049) and CYP3A4*1B_-392A/G (rs2740574), while only one patient homozygous for the variant genotype of CYP3A5*3_6986A/G (rs776746) was found.

5.4 RESULTS OF SNPs VALIDATION

Due to technical problem, the validation was performed only on 12 out of 15 polymorphisms. As mentioned above, in order to validate the analytical method used, 12 SNPs object of this study have been analyzed with two different techniques (Pyrosequencing or TaqMan® Assay and Illumina GoldenGate Assay). The analyses were conducted in duplicate for about 250 cases with a 100% accordance between the results obtained, demonstrating the validity of the analytical techniques employed.

In particular, besides of being analyzed with the classical methods of allelic discrimination by TaqMan® probes or Pyrosequencing, the CYP19A1_Ex11+410A/C (rs4646), CYP19A1_47T/C (rs700519), CYP19A1_1558T/C (rs10046), CYP3A4*1B_-392A/G (rs2740574), CYP3A5*3_6986A/G (rs776746), RIZ1_delP704 (rs2308040), COMT_12A/G (rs4680), ESR1_497T/C (rs2234693), ESR2_1082A/G (rs1256049), ESR2_1730A/G (rs4986938), CYP17A1_27A/G (rs743572) and CYP1B1*3_4326G/C (rs1056836) SNPs were also tested with the Illumina GoldenGate Assay.

The validated polymorphisms, the technique used, the number of samples genotyped in parallel, the number of discordant samples and the % of agreement between the methods, are reported in Table 21.

Polymorphism	Method of validation	N° samples	N° discordant samples	% accordance
CYP19A1_ex11+410A/C (rs4646)	TaqMan®	257	0	100%
CYP3A5*3_6986A/G (rs776746)	TaqMan®	247	0	100%
COMT_12A/G (rs4680)	TaqMan®	257	0	100%
ESR1_497T/C (rs2234693)	TaqMan®	252	0	100%
ESR2_1082A/G (rs1256049)	TaqMan®	253	0	100%
ESR2_1730A/G (rs4986938)	TaqMan®	257	0	100%
CYP17A1_27A/G (rs743572)	TaqMan®	257	0	100%
CYP19A1_47T/C (rs700519)	Pyrosequencing	257	0	100%
CYP19A1_1558T/C (rs10046)	Pyrosequencing	257	0	100%
CYP3A4*1B_-392A/G (rs2740574)	Pyrosequencing	257	0	100%
RIZ1_delP704 (rs2308040)	Pyrosequencing	252	0	100%
CYP1B1*3_4326G/C (rs1056836)	TaqMan®	125	0	100%

Table 21. List of validated SNPs, the method, the number of samples used for the validation.

The results regarding the reproducibility of the data obtained were fully satisfactory, ensuring the quality of the analysis of the entire set of polymorphisms analyzed.

5.5 CLINICAL DATA MANAGEMENT

Clinical data collection was a key process in the management of the study due to its importance in ensuring high-quality and reliable results.

According to the aims of the study, the specific case report forms (CRFs) reported below were created and delivered to the responsible clinicians of each enrolling center.

1. Patient registration form (Form 1): used for the collection of patients' personal data and for eligibility assessment
2. Primitive tumor form (Form 2): used to delineate the tumor characteristics (date of diagnosis, stage, metastatic tumor sites) and the therapeutic course that patients underwent before exemestane treatment
3. Follow up form (Form 3):to register every 3 months the patients' status (dead or alive)
4. Summary form – hormone therapy (Form 4): to collect the starting and the ending date of exemestane therapy and the cause of treatment interruption
5. Toxicity form (Form 5):to collect the toxicities developed during the treatment. This form needed to be filled out in correspondence of each clinical evaluation.
6. Response evaluation form (Form 6):for response assessment in the same metastatic site patients had at enrollment. This form needed to be filled out in correspondence of each clinical evaluation.

The CRFs were accompanied with a detailed operating instructions form in order to support their compiling.

Once completed, the CRFs were sent back to our center and the data collected were entered in a purposely-built database.

The database was an Excel file with different sheets absolving different proposes:

- Genetic data: this sheet contained a list of the polymorphisms analyzed with the related genotype for each patient enrolled;
- CRF 1-4: this sheet was appositely built for containing the data reported in the first four CRFs, that is the registration, primitive tumor, follow-up and hormone therapy data;
- Toxicity sheets: there were many toxicity sheet dedicated to the collection of safety data every 8 weeks, accordingly to the timing imposed by the protocol (an efficacy and safety assessment were requested every 8 weeks of treatment): 1-8 weeks; 9-16 week; 17-24 weeks; 25-32 weeks; 33-40 weeks; 41-48 weeks; 49-56 weeks; 57-64 weeks; 65-72 weeks; 73-80 weeks; over 80 weeks. These sheet contained a the same list of toxicities reported in the toxicity form of CRFs and, for each event, the starting and ending date, the grade, and the relationship with the drug were recorded.

- Response: as for the toxicity evaluation, the response sheet was divided into sections corresponding to 8 weeks each. Each section had to be completed with the date of the examination, the size and the response to treatment of the target and non target lesions assessed during the each evaluation.

During the data entry, their consistency was assessed in order to identify ineligible patients or data which needed a revision.

If any nonconformity was found the physician was requested to revise and provide the correct data.

Finally, when all the available data were collected, a process of conformation was done in order to make the data more homogeneous. These data were subsequently used for the statistical association.

FORM 1

PATIENT REGISTRATION FORM

To be forwarded via fax to: 0434 XXXXX

or via e-mail to the following address: XXXXXX@cro.it

Patient information **Clinical file No.: _____**

Initials □□□□ **Date of birth** □□ □□ □□

Center No. [ ]

Physician's name **Fax** __/__/__/__/__/__/__/__/__/__

_____

Physician's signature:

INCLUSION CRITERIA	**YES**		**NO**
1) Written informed consent			
2) Histological diagnosis of metastatic breast cancer			
3) ≥ 55 years of age (post-menopausal women)			
4) Performance status 0-2 (WHO sec.)			
5) At least one measurable lesion (one- or two-dimensional)			
6) No target lesions as defined by RECIST criteria			
7) Neutrophils ≥ 1500μL, platelets ≥ 100,000μL Hb ≥ 9 g/dl			
8) Life expectancy > 3 months			

EXCLUSION CRITERIA	**YES**		**NO**
1) Pregnancy, lactation, unsatisfactory contraception			
2) Prior exemestane treatment			
3) Presence of multiple symptomatic cerebral metastases			
4) Other serious concomitant pathologies			
5) Serious functional alteration of visceral and metabolic disease			
6) Radiotherapy or major surgery within 4 weeks from beginning of exemestane therapy			
7) Prior or concomitant neoplasia (except for *in situ* cervical cancer)			
8) Concomitant treatment with another antineoplastic therapy			

The following space must be filled in by the coordinating center

Patient code □□□ **Registration date** □□ □□ □□

FORM 2 (I)

PRIMITIVE TUMOR FORM	**Center [ ]**	**Patient code** □□□
Date of diagnosis	____/____/____	
Diagnosed disease stage (TNM- Tumor Node Metastasis)	[I]	[II]
	[III]	[IV]
Menopause		
Physiological	Date of the last menstruation ____/____/____	
From chemotherapy	Date of the last menstruation ____/____/____	
From LHRH analogues	Date of the last menstruation ____/____/____	
Ovariectomy	Date of the last menstruation ____/____/____	
Isterectomy without ovariectomy in patients > 50 years old	Date of the last menstruation ____/____/____	

Disease sites upon recruitment	**Measurement of the disease site (one or two dimensional)**	**Examination used for the measurement**	**Examination date**
Target lesions (RECIST's criteria)			
1) _____	cm ..... x cm……	[ ] _____	____/____/____
2) _____	cm ..... x cm……	[ ] _____	____/____/____
3) _____	cm ..... x cm……	[ ] _____	____/____/____
4) _____	cm ..... x cm……	[ ] _____	____/____/____
Non target lesions (RECIST's criteria)			
1) _____	cm ..... x cm……	[ ] _____	____/____/____
2) _____	cm ..... x cm……	[ ] _____	____/____/____
3) _____	cm ..... x cm……	[ ] _____	____/____/____

**for each disease site, indicate a maximum of 3 valuable lesions*						
^legend for the examinations that can be used for evaluation of the response:						
[1] CT	*[2] Radiography*	*[3] Echography*	*[4] Scintigraphy*	*[5] NMR*	*[6] OE (objective examination)*	*[7] other*

FORM2 (II)

PRIMITIVE TUMOR FORM		Center [ ]	Patient code □□□
Receptor Status			
ER	[POS]	[NEG]	%_____
PgR	[POS]	[NEG]	%_____
HER2	[POS] 3+ or amplified	[NEG]	%_____
KI67	[HIGH] (≥15%)	[LOW]	%_____

PRIMITIVE TUMOR SURGERY	[YES]	[NO]
Kind of the surgical intervention	Date of the surgical intervention	
_____	___/___/___	
_____	___/___/___	

RADIOTHERAPY (adjuvant)	[YES]	[NO]
Sites of irradiation	Start date	Stop date
_____	___/___/___	___/___/___
_____	___/___/___	___/___/___

CHEMOTHERAPY (neoadjuvant, adjuvant or metastatic)		[YES]	[NO]
Regimen (drugs)	N° of cycles	Start date	Stop date
Neoadjuvant:_____	_____	___/___/___	___/___/___
Adjuvant:_____	_____	___/___/___	___/___/___
Metastatic:_____	_____	___/___/___	___/___/___

HORMONE THERAPY (adjuvant or metastatic)		[YES]	[NO]
Drugs	Dose mg/die	Start date	Stop date
Adjuvant:_____	_____	___/___/___	___/___/___
Metastatic:_____	_____	___/___/___	___/___/___

FORM 3

FOLLOW–UP FORM *(to be filled in every 3 months after the end of treatment)*	**Center** **[ ]**	**Patient code** □□□

Progression since first line	**[1] NO**	**[2] YES**
Date of the first progression	**____/____/____**	
Site of the first progression	**_____**	
Treatment received for the first progression	**_____**	
Patient status ***(3 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(6 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(9 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(12 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(15 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(18 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(21 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	
Patient status ***(24 months)***	**[1] alive**	**[2] dead**
Date of the visit or death	**____/____/____**	

FORM 4

SUMMARY FORM – HORMONE THERAPY	**Center** **[ ]**	**Patient code** □□□

Date of inclusion in the study ____/____/____

Date of the first administration ____/____/____

Date of the last administration ____/____/____

Reason for interruption of the treatment

[1] end of the scheduled cycles

[2] progression of the disease

[3] toxicity

[4] death during treatment

[5] rejection by the patient

Best response obtained

[1] CR [2] PR

[3] SD [4] PD

This form must be filled in by the researcher,
only after the conclusion of the therapy

FORM 5 (I)

TOXICITY FORM (N°...evaluation) *(see Appendix II to the protocol for the NCI–CTC criteria)*	Center [ ]	Patient code □□□

WEEKS 1-8	YES	NO	Start date:	Stop date:	Grade *	Relationship ^
Blood/ bone marrow						
Hemoglobin			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Leukocytes			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Neutrophils			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Platelets			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Transfusion			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other _____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Cardiovascular						
Cardiovascular disease (no myocardial infarction)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Thromboembolic disease			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Cardiac Ischemia /infarction			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Bleeding						
Hemorrhage/bleeding with grade 3 or 4 thrombocytopenia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Hemorrhage/bleeding without grade 3 or 4 thrombocytopenia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Vaginal bleeding			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other _____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Constitutional symptoms						
Depression			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Insomnia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Dizziness			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Sweating			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Hot flashes			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Fatigue			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Weight loss			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Cramps			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other _____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5

*** Grade**

1. MILD
2. MODERATE
3. SEVERE
4. LIFE-THREATENING
5. DEATH

^ Relationship

1. The adverse event is NOT related to the drug
2. It's NOT CLEAR whether the adverse event is related to the drug
3. The adverse event MAY be related to the drug
4. The adverse event is PROBABLY related to the drug administration
5. The adverse event is CLEARLY related to the drug administration

FORM 5 (II)

TOXICITY FORM (N°... evaluation) *(see Appendix II to the protocol for the NCI–CTC criteria)*					Center [ ]	Patient code □□□
WEEK:_____	**YES**	**NO**	**Start date**	**Stop date**	**Grade**	**Relationship ^**
Dermatology						
Alopecia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Rash/desquamation			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Hand and foot syndrome			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Articulations						
Osteoporosis			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Arthralgia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Gastrointestinal						
Anorexia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Dehydratation			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Diarrhea			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Epigastralgia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Nausea			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Stomatitis			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Vomiting			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Hepatic						
Bilirubin			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
GGT			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
SGOT (AST)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
SGPT (ALT)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5

*** Grade**

1. MILD
2. MODERATE
3. SEVERE
4. LIFE-THREATENING
5. DEATH

^ Relationship

1. The adverse event is NOT related to the drug
2. It's NOT CLEAR whether the adverse event is related to the drug
3. The adverse event MAY be related to the drug
4. The adverse event is PROBABLY related to the drug administration
5. The adverse event is CLEARLY related to the drug administration

FORM 5 (III)

TOXICITY FORM (N°... evaluation) ***(see Appendix II to the protocol for the NCI–CTC criteria)***	**Center** [ ]	**Patient code** □□□

WEEKS 1-8	**YES**	**NO**	**Start date:**	**Stop date:**	**Grade ***	**Relationship ^**
Infection/ neutropenic fever						
Catheter-related infection			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Febrile neutropenia (ANC<1.0 X 10^9/L fever >38.5°C)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Infection with grade 3 or 4 neutropenia (ANC<1.0 X 10^9/L)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Infection without neutropenia			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Ocular/visual						
Visual disturbances (_____)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Visual disturbances (_____)			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Pain						
Pain or aches			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Headaches			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Respiratory						
_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Renal/ genitourinary						
Creatinine			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Gynecologic symptoms _____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other _____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
Other						
_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5
_____			___/___/___	___/___/___	1 2 3 4 5	1 2 3 4 5

*** Grade**

6. MILD
7. MODERATE
8. SEVERE
9. LIFE-THREATENING
10. DEATH

^ Relationship

6. The adverse event is NOT related to the drug
7. It's NOT CLEAR whether the adverse event is related to the drug
8. The adverse event MAY be related to the drug
9. The adverse event is PROBABLY related to the drug administration
10. The adverse event is CLEARLY related to the drug administration

FORM 6 (I)

RESPONSE EVALUATION *(after 8 weeks)*	**Center [ ]**	**Patient code □□□**

Sites of the disease upon recruitment*	**Measurement of the disease sites *(one- or two-dimensional)***		**Examination used^**	**Examination date**	**Response #**
Target lesions (RECIST's criteria)*					
1) _____	cm ..... x cm	not detectable	[ ] _____	____/____/____	[ ]
2) _____	cm ..... x cm	not detectable	[ ] _____	____/____/____	[ ]
3) _____	cm ..... x cm	not detectable	[ ] _____	____/____/____	[ ]
New site_____	cm ..... x cm	not detectable	[ ] _____	____/____/____	[ ]
Non target lesions (RECIST's criteria)*					
1) _____			[ ] _____	____/____/____	[ ]
2) _____			[ ] _____	____/____/____	[ ]
3) _____			[ ] _____	____/____/____	[ ]
New site _____			[ ] _____	____/____/____	[ ]
Overall response					**[ ]**

**** for each disease site, indicate a maximum of 3 valuable lesions***		***^legend for the examinations that can be used for evaluation of the response***	
[1] CT	*[3] Echography*	*[5] NMR*	*[7] Other (specify the type of the examination)*
[2] Radiography	*[4] Scintigraphy*	*[6] OE (objective examination)*	
# CR = complete response	*PR = partial response*	*SD = stabilization of the disease*	*PD = progression of the disease*

5.6 RELATIONSHIP BETWEEN POLYMORPHISM AND RESPONSE TO TREATMENT

Association between genotypes and clinical response of each patient was assessed by the two-sided Fisher's exact test.

To do so, the best clinical response obtained after at least 8 weeks of treatment was the parameter employed for the statistical association with the polymorphisms genotypes.

The possible best clinical response were: CR, PR, SD and PD. Patients were grouped into two sets, according to the type of clinical response obtained on the basis of the criteria above described for the definition of RR and CB.

Among the 15 investigated polymorphisms, CYP1B1*3_4326G/C (rs1056836) resulted significantly associated with the response to the drug. No other associations with the clinical outcome were found for the other SNPs or STRs (data not shown).

Here below are reported, besides the results of CYP1B1*3_4326G/C (rs1056836), only the statistical analyses of CYP19A1_Ex11_410A/C (rs4646) SNP, which was the primary genetic end point of the clinical study.

- ✓ **CYP1B1*3_4326G/C (rs1056836)**

The only polymorphism, among the 15 analyzed, to be significantly associated with the response to treatment in terms of RR was the CYP1B1*3_4326G/C (rs1056836) SNP. For this polymorphism 275 patients were analyzed. In Table 22 is shown the distribution of the best clinical responses obtained from patients stratified by genotype.

Genotypes	CR	% CR	PR	% PR	SD	% SD	PD	% PD	Total
CC	3	3%	17	19%	48	54%	21	24%	89
GC	10	8%	38	30%	57	45%	22	17%	127
GG	7	12%	20	34%	22	37%	10	17%	59
GC+GG	17	9%	58	31%	79	42%	32	17%	186
CC+GC	13	6%	55	25%	105	49%	43	20%	216
GG	7	12%	20	34%	22	37%	10	17%	59
All	20	7%	75	27%	127	46%	53	19%	275

Table 22. Distribution of patients' treatment responses depending on the genotype of the CYP1B1*3_4326G/C SNP. CR = complete response, PR = partial response, SD = stable disease, PD = progression disease.

Regarding the Response Rate (RR = CR+PR), as seen from the Table 23, patients carrying the wild type CC genotype were the 22%, those with GC genotype were the 38% and those with the GG genotype were the 46%. In the second group (SD+PD), associated with a worse response to treatment, patients with the

CC genotype were the 78%, patients heterozygous were the 62% and those with the GG genotype were 54%.

Statistical analysis carried out to determine a possible association between genotype and response, showed that heterozygous subjects had a 2.10- fold higher probability then wild type subjects to be responsive to treatment (OR = 2.10; 95% CI = 3.85 to 1.14; p = 0.0179). An even stronger association was found when patients carrying the GG variant genotype were compared with the wild type ones with an OR = 2.91; 95% CI 5.88 – 1.25 and p = 0.0039. This result was confirmed by either the dominant and the recessive models in which subjects carrying at least one mutated allele were found to have a higher probability of responding to therapy (dominant model: OR = 2.33; 95% CI = 4.17 to 1.31, p = 0.0043; recessive model: OR = 1.83; 95% CI 3.33 to 1.02; p = 0.0456).

Genotypes	CR+PR	%	SD+PD	%	Tot		OR	95% CI	P	χ^2	P
CC	20	22%	69	78%	89		1				
GC	48	38%	79	62%	127		2.10	3.85 – 1.14	0.0179		
GG	27	46%	32	54%	59		2.91	5.88 – 1.25	0.0039	9.61	0.0082
GC+GG	75	40%	111	60%	186	DOM	2.33	4.17 – 1.31	0.0043		
CC+GC	68	31%	148	69%	216		1				
GG	27	46%	32	54%	59	REC	1.83	3.33 – 1.02	0.0456		
CC	95	35%	180	65%	275						

Table 23. RR evaluation depending on the genotype of the CYP1B1*3_4326G/C SNP. CR = complete response, PR = partial response, SD = stable disease, PD = progression disease; DOM = dominant model, REC = recessive model.

Regarding the Clinical Benefit (CB = CR + PR + SD) assessment, no statistically significant associations were found, as can be seen in the Table 24.

Genotypes	CR+PR+SD	%	PD	%	Tot		OR	95% CI	P	χ^2	P
CC	68	76%	21	24%	89		1				
GC	105	83%	22	17%	127		1.47	2.86 – 0.75	0.2999		
GG	49	83%	10	17%	59		1.52	3.45 – 0.65	0.4109	1.58	0.4529
GC+GG	154	83%	32	17%	186	DOM	1.49	2.78 – 0.80	0.2526		
CC+GC	173	80%	43	20%	216		1				
GG	49	83%	10	17%	59	REC	1.22	2.63 – 0.57	0.7113		
CC	222	81%	53	19%	275						

Table 24. CB evaluation depending on the genotype of the CYP1B1*3_4326G/C SNP. CR = complete response, PR = partial response, SD = stable disease, PD = progression disease; DOM = dominant model, REC = recessive model.

✓ **CYP19A1_Ex11_410A/C (rs4646)**

Genotypes	CR	% CR	PR	% PR	SD	% SD	PD	% PD	Total
CC	8	6%	35	24%	73	51%	28	19%	144
AC	8	7%	34	31%	44	41%	22	20%	108
AA	4	17%	6	26%	10	43%	3	13%	23
AC+AA	12	9%	40	31%	54	41%	25	19%	131
CC+AC	16	6%	69	27%	117	46%	50	20%	252
AA	4	17%	6	26%	10	43%	3	13%	23
All	20	7%	75	27%	127	46%	53	19%	275

Table 25. Distribution of patients' treatment responses depending on the genotype of the CYP19A1_Ex11_410A/C SNP. CR = complete response, PR = partial response, SD = stable disease, PD = progression disease.

No significant trend of association was highlighted between the polymorphism and clinical response obtained, even using the dominant or recessive genetic models, neither for the RR (CR+PR vs SD+PD) (Tab. 26) or for the CB (Tab. 27).

Genotypes	CR+PR	%	SD+PD	%	Tot		OR	95% CI	P	χ^2	P
CC	43	30%	101	70%	144		1				
AC	42	39%	66	61%	108		1.49	2.50 – 0.88	0.1410		
AA	10	43%	13	57%	23		1.82	4.35 – 0.74	0.2292	3.11	0.2112
AC+AA	52	40%	79	60%	131	DOM	1.55	2.56 – 0.93	0.0993		
CC+AC	85	34%	167	66%	252		1				
AA	10	43%	13	57%	23	REC	1.51	3.57 – 0.64	0.3654		
All	95	35%	180	65%	275						

Table 26. RR evaluation depending on the genotype of the CYP19A1_Ex11_410A/C SNP. CR = complete response, PR = partial response, SD = stable disease, PD = progression disease; DOM = dominant model, REC = recessive model.

Genotypes	CR+PR+SD	%	PD	%	Tot		OR	95% CI	P	χ^2	P
CC	116	81%	28	19%	144		1				
AC	86	80%	22	20%	108		0.94	1.75 – 0.51	0.8743		
AA	20	87%	3	13%	23		1.61	5.88 – 0.45	0.5747	0.66	0.7189
AC+AA	106	81%	25	19%	131	DOM	1.02	1.85 – 0.56	1.0000		
CC+AC	202	80%	50	20%	252		1				
AA	20	87%	3	13%	23	REC	1.65	5.88 – 0.47	0.5847		
All	222	81%	53	19%	275						

Table 27. CB evaluation depending on the genotype of the CYP19A1_Ex11_410A/C SNP. CR = complete response, PR = partial response, SD = stable disease, PD = progression disease; DOM = dominant model, REC = recessive model.

5.7 RELATIONSHIP BETWEEN POLYMORPHISMS AND TIME TO PROGRESSION

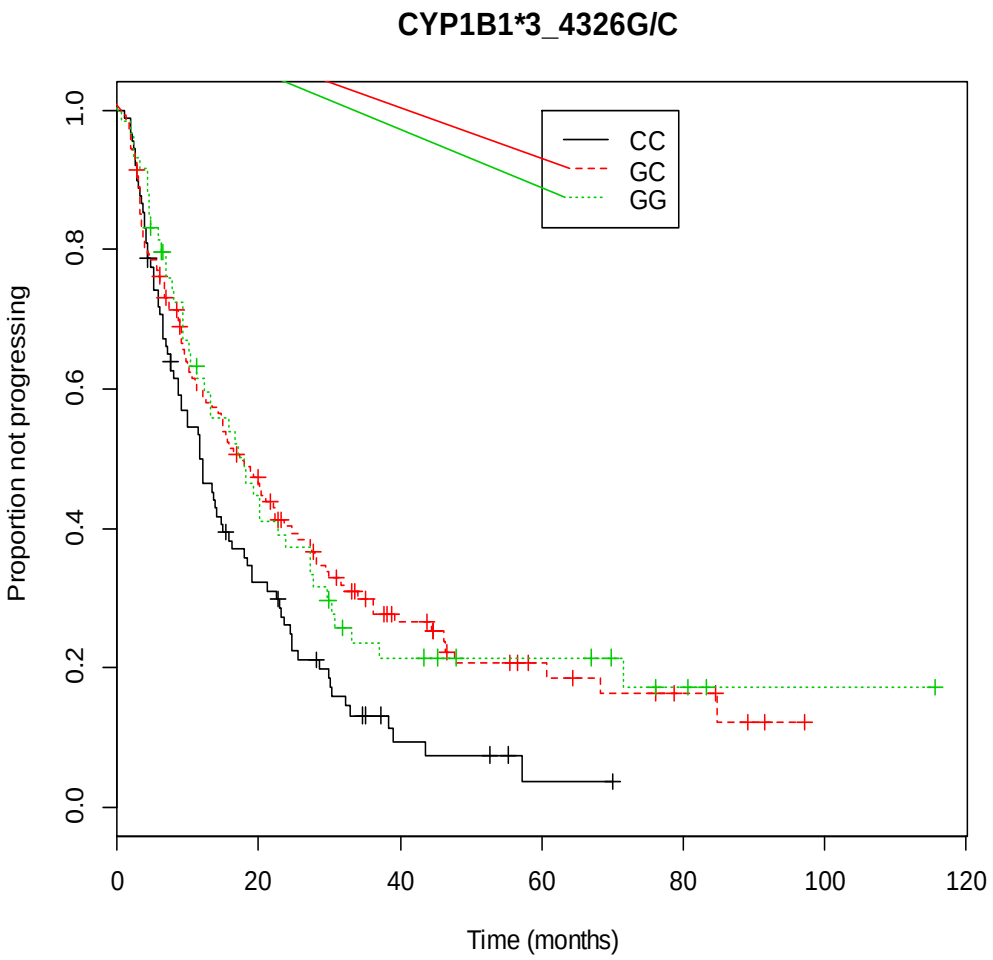
The Kaplan-Meier product-limit method and the log-rank test statistic were applied in order to evaluate the association between TTP and the 15 polymorphisms analyzed.

Only the CYP1B1*3_4326G/C (rs1056836) and the CYP19A1_1558T/C (rs10046) SNP affected the TTP in a statistically significant manner. Results are reported as Hazard Ratios (HR) and the relative 95% CI and only the data relative to the two SNPs significantly associated with the TTP and the CYP19A1_Ex11_410A/C SNP (rs4646) (primary end-point of the clinical study) are shown below. On the Kaplan-Meier curves, the abscissa indicates the time expressed in months at which the event (progression) occurred, while in the ordinate axis there is the proportion of not progressing patients.

Median follow-up of patients was 35 months (2-153 months).

- ✓ **CYP1B1*3_4326G/C (rs1056836)**

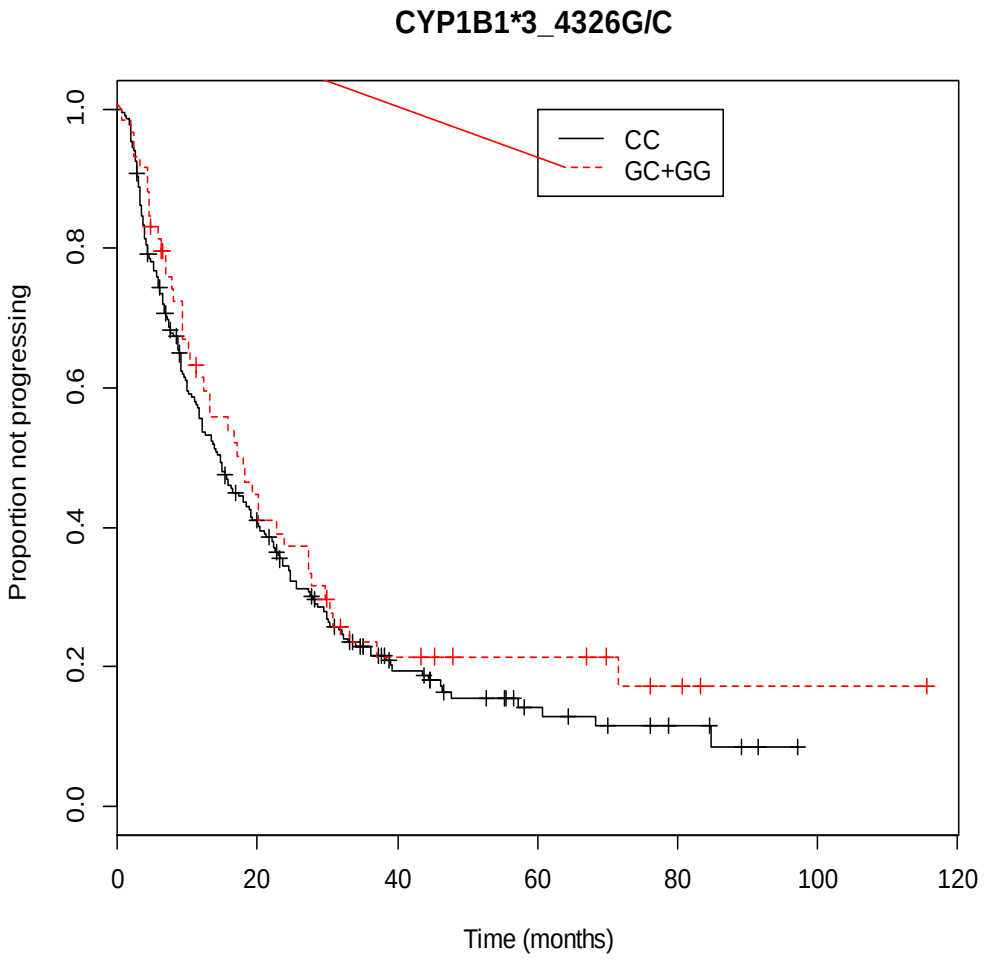
This SNP, which in this study was found to be linked with the clinical outcome of patients, was statistically associated with the TTP too. Progression occurred in 216 patients: 89 of them had a CC genotype, 94 were heterozygous while 59 carried the homozygous mutated GG genotype. Median TTP increased in a statistically significant manner as the number of mutated alleles increased. Indeed, patients with the CC genotype, which was taken as the reference category, had a median TTP of 11.8 months, while heterozygous patients had a median TTP of 17.3 months, with an HR of progressing 0.66-fold lower than the reference category's one (95% CI 0.49 – 0.89, p = 0.0073). Patients with the homozygous mutated genotype had the higher median TTP and its value was 18.0 months (HR = 0.66; 95% CI 0.45 – 0.96; p = 0.0280) (Fig. 35).



CYP1B1*3_4326G/C	N° cases (%)	N° events (%)	Median	p-value log-rank	HR	95%IC	p-value
CC	89 (32%)	78 (87%)	11.8	0.0141	Reference		
GC	127 (46%)	94 (74%)	17.3		0.66	0.49 – 0.89	0.0073
GG	59 (21%)	44 (74%)	18.0		0.66	0.45 – 0.96	0.0280

Figure 35. TTP according to the CYP1B1*3_4326G/C genotypes. In red are highlighted the statistically significant log-rank and the p-values.

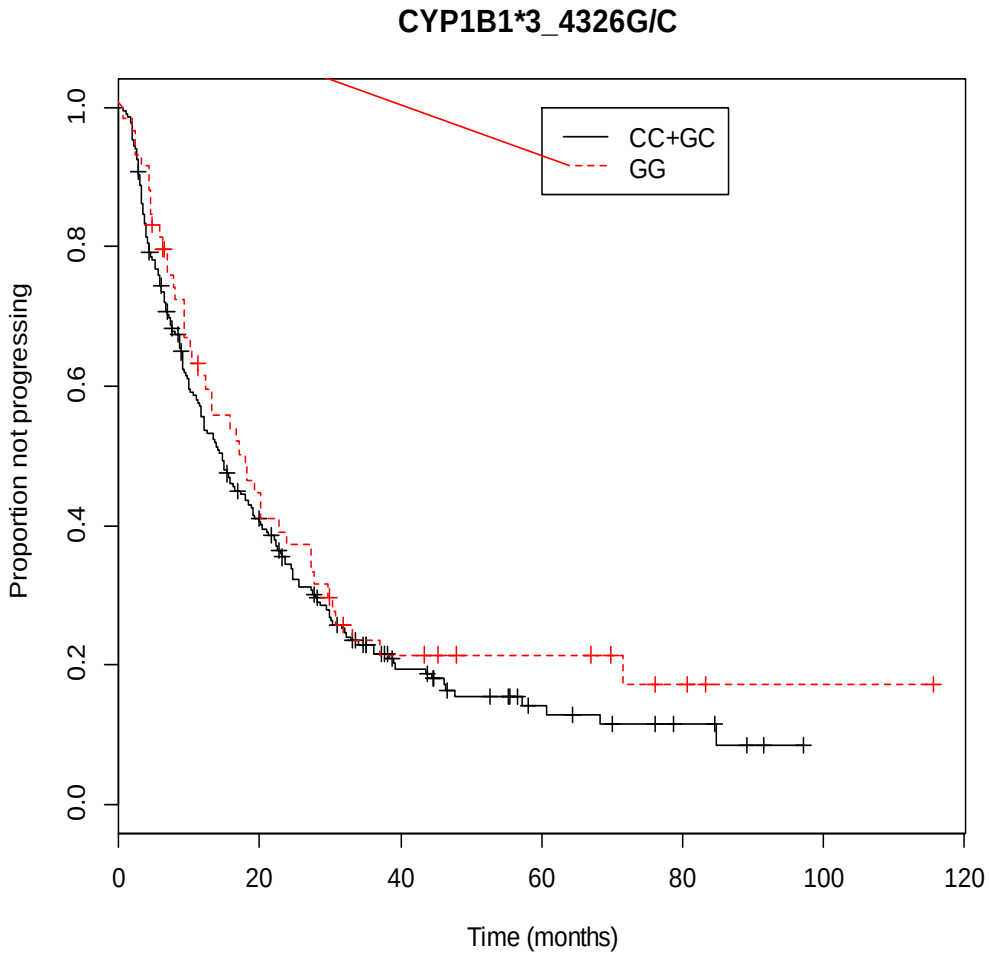
This result was confirmed when the dominant statistical model was applied. Indeed, patients carrying at least one mutated allele (CG+GG group) had a statistically significant longer median TTP (17.3 months) with respect to the reference category (11.8 months) represented by the CC genotype patients (HR = 0.66; 0.50-0.87; p = 0.0037) (Fig. 36).



CYP1B1*3_4326G/C	N° cases (%)	N° events (%)	Median	p-value log-rank	HR	95%IC	p-value
DOMINANT MODEL							
CC	89 (32%)	78 (88%)	11.8	0.0035	Reference		
CG+GG	186 (68%)	138 (71%)	17.3		0.66	0.50 – 0.87	0.0037

Figure 36. TTP according to the dominant model for the CYP1B1*3_4326G/C SNP. In red are highlighted the statistically significant log-rank and p-value for the comparison between the CC group and patients carrying at least one G allele.

On the contrary, no statistically significant association was found when the recessive model was applied (Fig. 37).

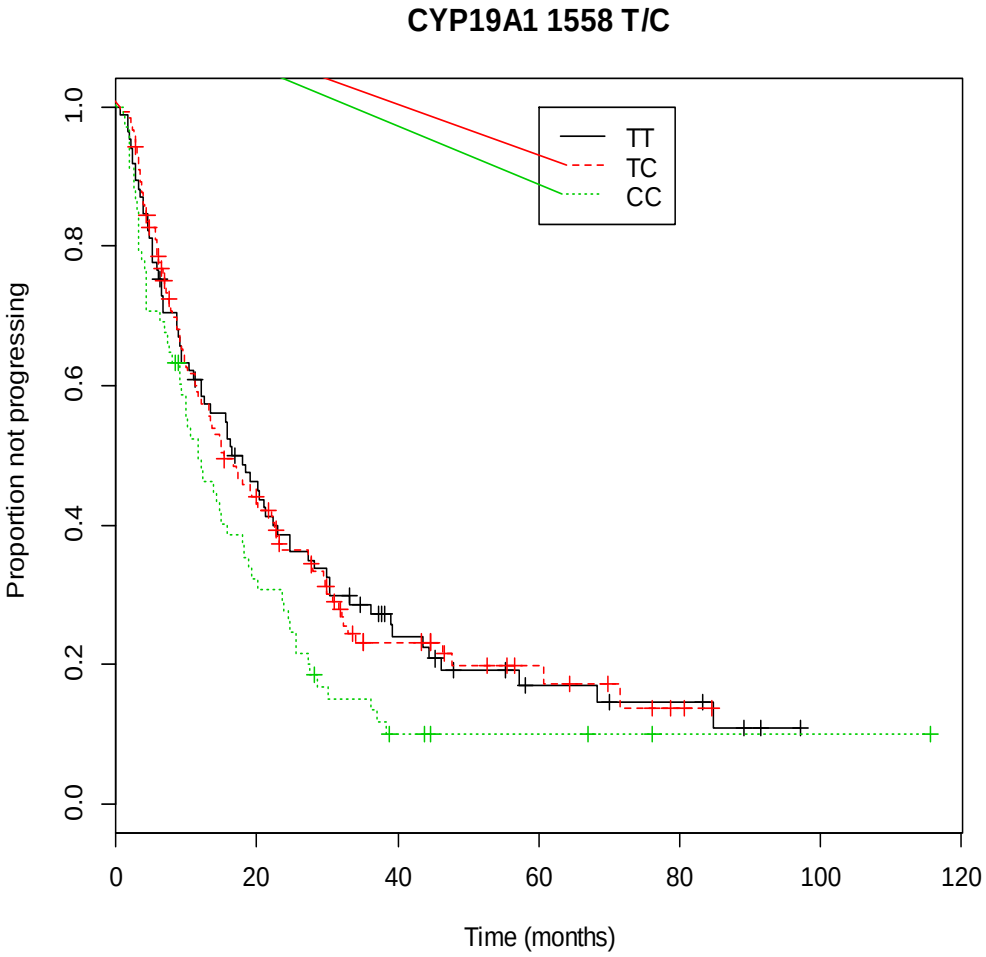


CYP1B1*3_4326G/C	N° cases (%)	N° events (%)	Median	p-value log-rank	HR	95%IC	p-value
RECESSIVE MODEL							
CC+GC	216 (79%)	172 (80%)	14.8	0.325	Reference		
GG	59 (21%)	44 (75%)	18.0		0.847	0.61 – 1.18	0.33

Figure 37. TTP according to the recessive model for the CYP1B1*3_4326G/C SNP.

✓ **CYP19A1_1558T/C (rs10046)**

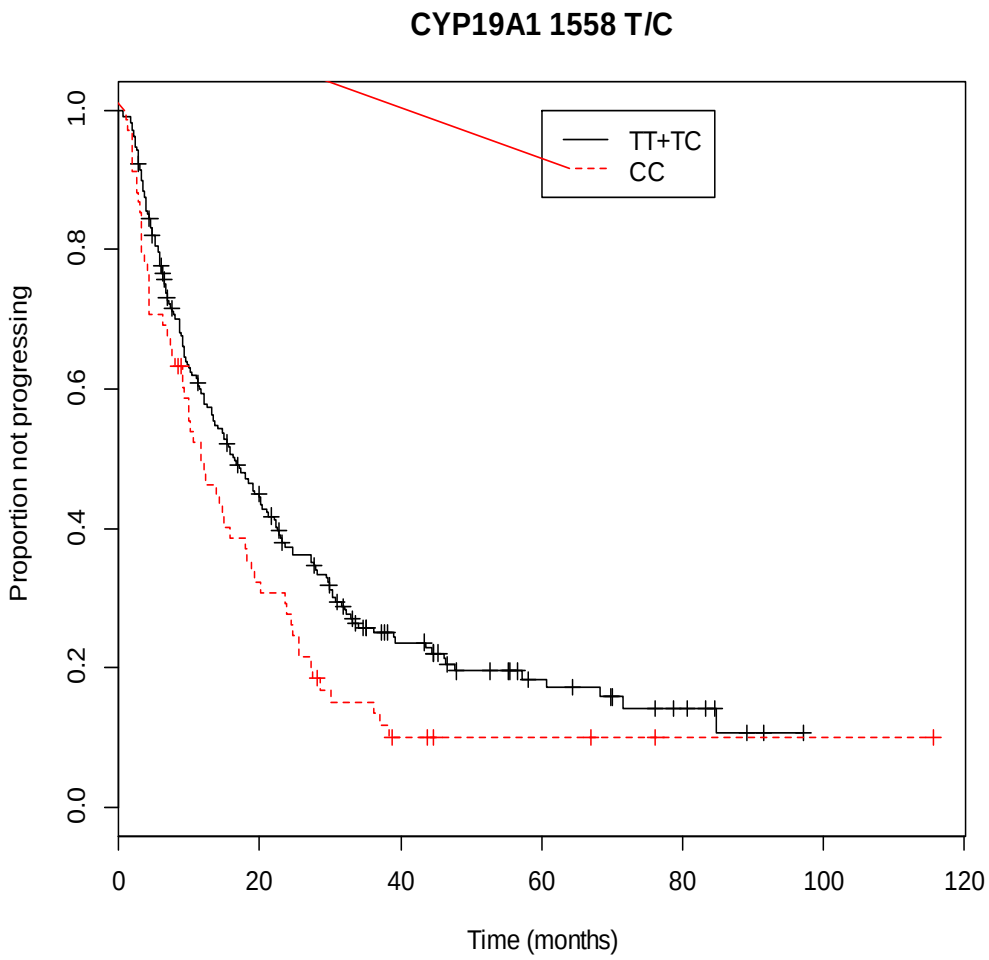
As regard this polymorphism a trend for association was found between wild type (TT) and mutated (CC) patients. In particular, the first group was associated with a longer TTP with a median value of 16.5 months, compared with a TTP of 11.6 months for the mutated patients (HR = 1.41, 95% CI 0.99 - 2.00, p = 0.056) (Fig. 38).



CYP19A1_1558T/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
TT	85	68	16.5	0.0881	Reference		
TC	122	89	15.4		1.01	0.73 – 1.39	0.950
CC	68	59	11.6		1.41	0.99 – 2.00	0.056

Figure 38. TTP according to the CYP19A1_1558T/C genotypes. In orange is highlighted the slightly significant p-value that compares the wild type (TT) group with the mutated (CC) one.

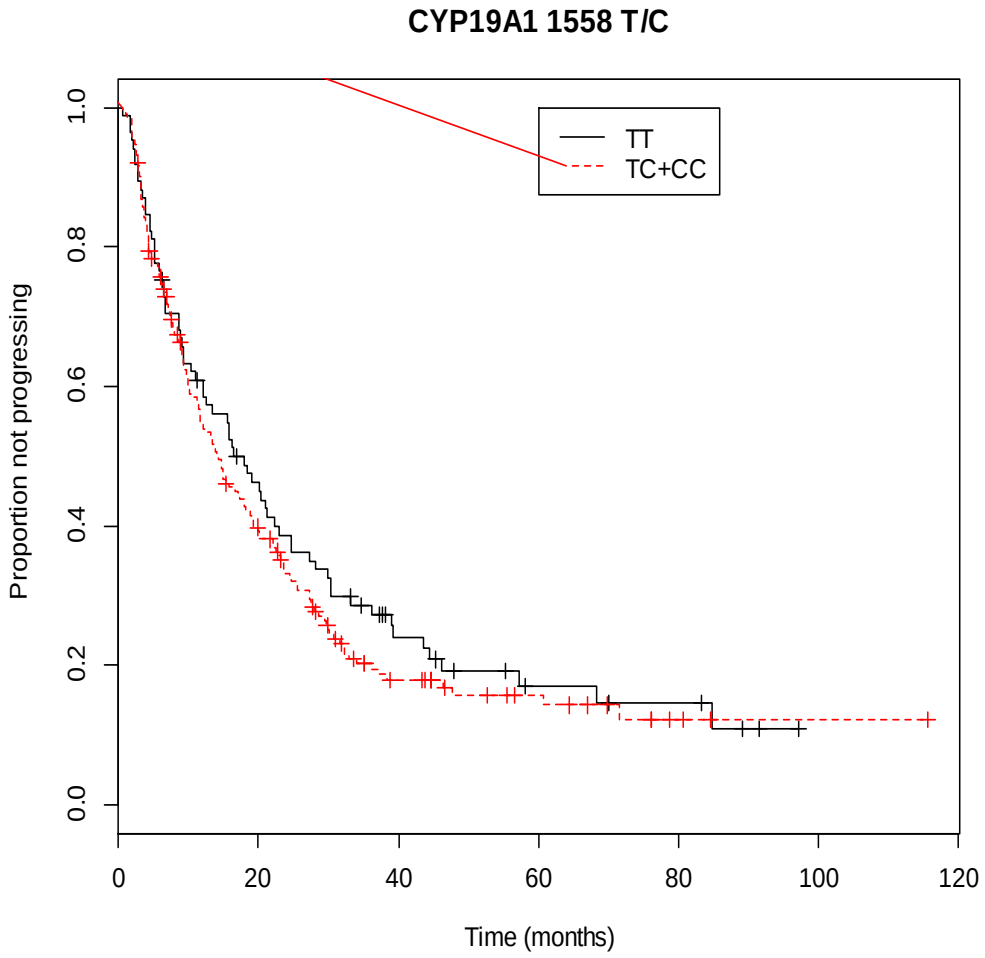
A little stronger and statistically significant association, instead, was found when the recessive model was applied. In particular, patients homozygous for the mutant allele (CC) had an HR of progressing 1.4-fold higher then patients carrying at least one wild type allele (95% CI = 1.04 – 1.89; p = 0.028) (Fig. 39).



CYP19A1_1558T/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
RECESSIVE MODEL							
TT+TC	207	157	16.5	0.0276	Reference		
CC	68	59	11.6		1.4	1.04 – 1.89	0.028

Figure 39. TTP according to the recessive model for the CYP19A1_1558T/C SNP. In red are highlighted the statistically significant log-rank and p-value for the comparison between the CC group and patients carrying at least one T allele.

On the contrary, no significant associations were found when the dominant model was applied (Fig. 40).

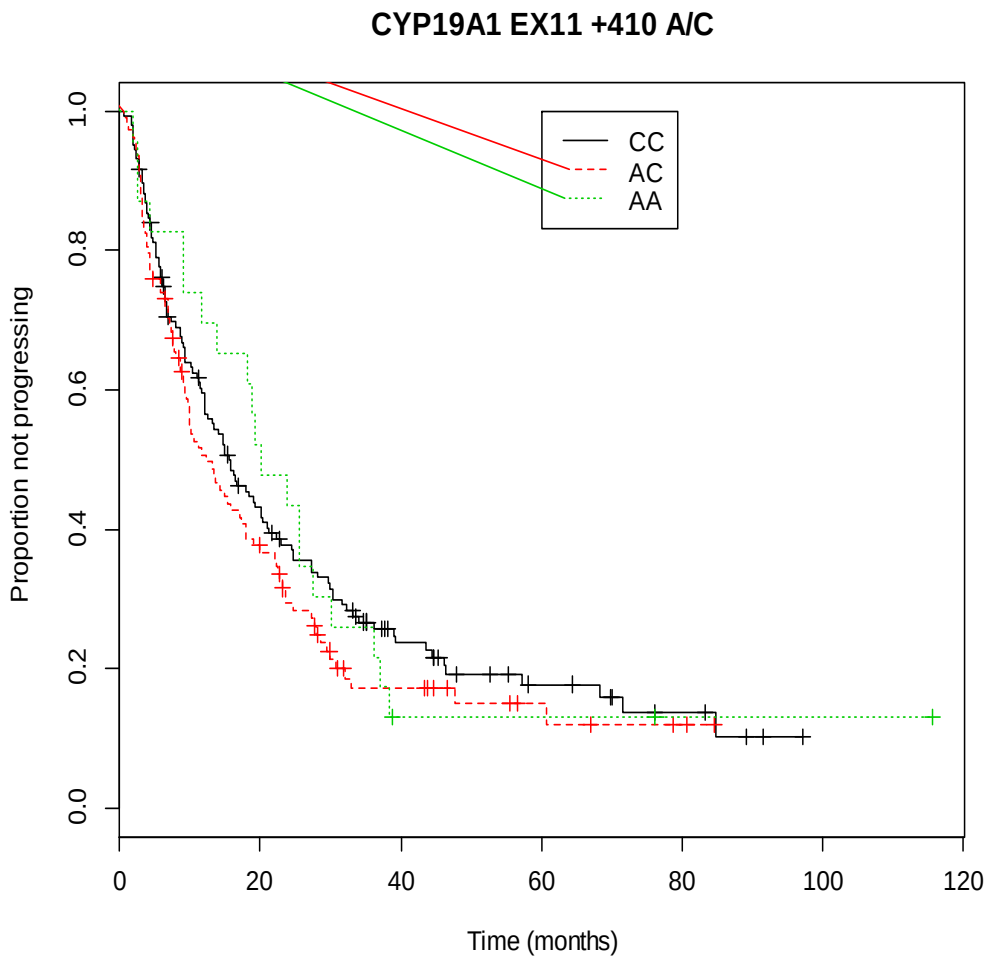


CYP19A1_1558T/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
DOMINANT MODEL							
TT	85	68	16.5	0.383	Reference		
TC+ CC	190	148	14.4		1.14	0.85 – 1.52	0.38

Figure 40. TTP according to the dominant model for the CYP19A1_1558T/C SNP.

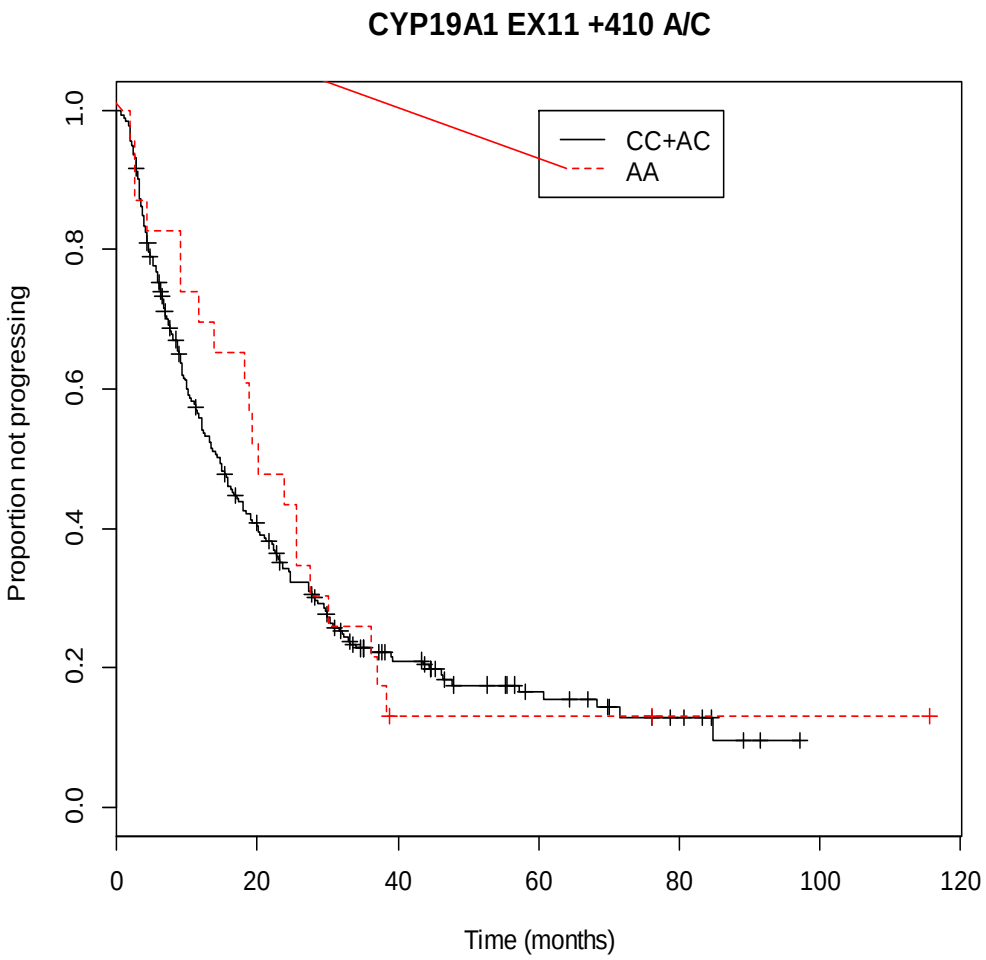
✓ **CYP19A1_Ex11_410A/C (rs4646)**

No statistically significant associations between CYP19A1_Ex11_410A/C (rs4646) genotypes and TTP were found, neither applying the dominant or recessive model, as can be seen by the Figures 41, 42 and 43 reported below.



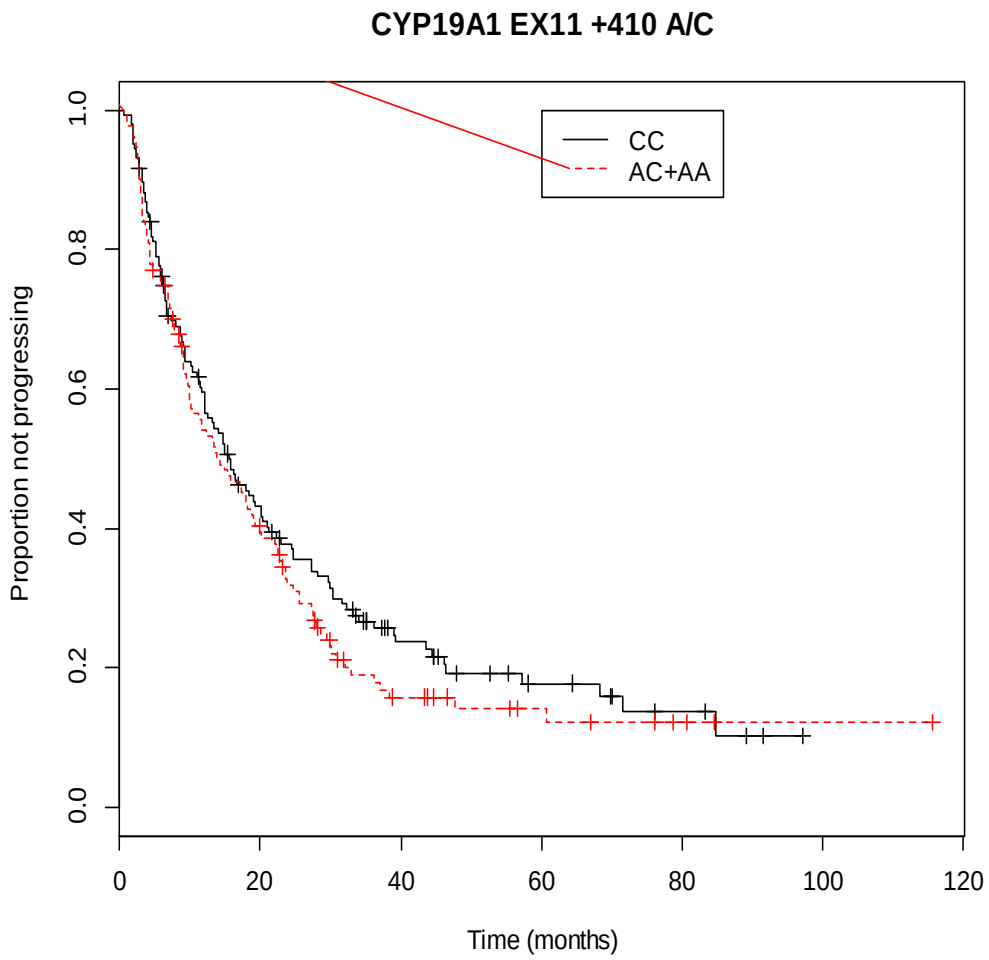
CYP19A1_EX11+410A/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
CC	144	111	15.5	0.414	Reference		
AC	108	85	12.3		1.198	0.90 – 1.58	0.21
AA	23	20	20.2		0.972	0.60 – 1.56	0.91

Figure 41. TTP according to the CYP19A1_Ex11+410A/C genotypes.



CYP19A1_EX11+410A/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
RECESSIVE MODEL							
CC+AC	252	196	14.8	0.667	Reference		
AA	23	23	20.2		0.903	0.68 – 1.20	0.67

Figure 42. TTP according to the recessive model for the CYP19A1_Ex11+410A/C SNP.



CYP19A1_EX11+410A/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
DOMINANT MODEL							
CC	144	111	15.5	0.314	Reference		
AC+AA	131	105	14.4		1.15	0.88 – 1.50	0.32

Figure 43. TTP according to the dominant model for the CYP19A1_Ex11+410A/C SNP.

5.8 RELATIONSHIP BETWEEN POLYMORPHISMS AND OVERALL SURVIVAL

Association between genotypes and OS was assessed for the 15 SNPs and STRs object of the study employing the Kaplan-Meier product-limit method and the log-rank test in order to evaluate if the SNPs investigated, and in particular the CYP1B1*3_4326G/C (rs1056836) SNP (resulted associated, in this cohort, with a better RR and a longer TTP) had an impact on patients' OS.

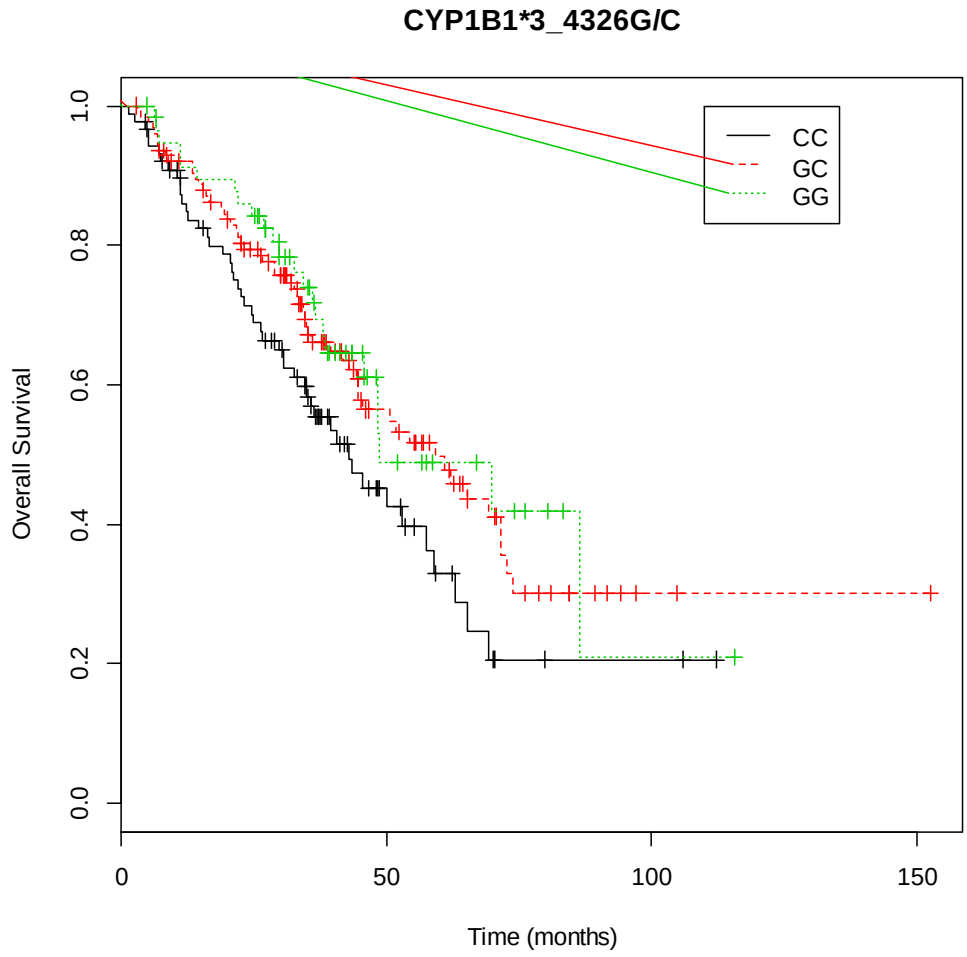
A statistically significant association between OS and genotypes was indeed found only for the CYP1B1*3_4326G/C (rs1056836) SNP, reinforcing its prognostic value.

As for the TTP analysis, here below are reported only the significant results and the data regarding the CYP19A1_Ex11_410A/C (rs4646) SNP, the primary end point of the clinical study.

Results are expressed as HR and the relative 95% CI. On the Kaplan-Meier curves, the abscissa indicates the time expressed in months at which the event (death or last follow up) occurred, while in the ordinate axis there is the percentage of surviving patients expressed as OS.

✓ **CYP1B1*3_4326G/C (rs1056836)**

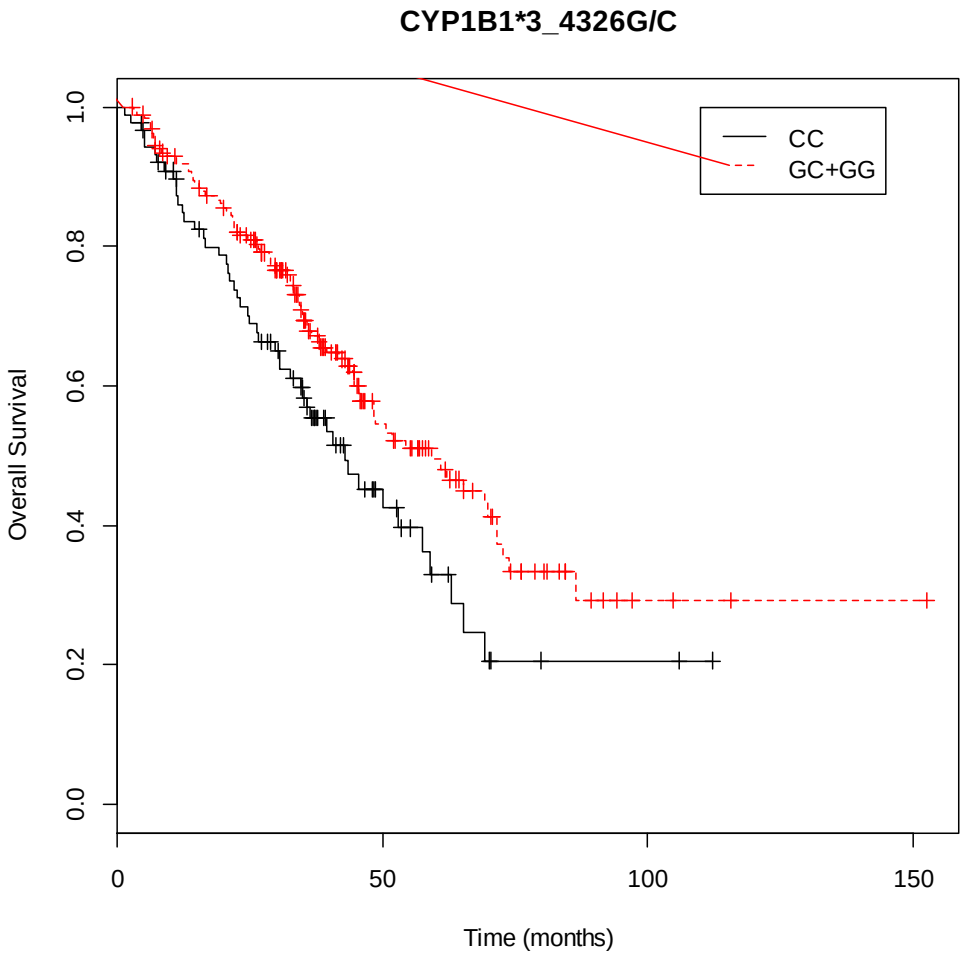
For this SNP a considerable trend toward significance was shown by the log-rank test (p = 0.0675), indicating that the difference seen in the OS among the three groups of patients was effectively affected by the genotypes. Indeed, heterozygous patients had a median OS of 59.4 months, while for the wild type (CC) group of patients, taken as reference category, it was of 43.0 months. This difference was just about statistically significant, with an HR of 0.682, a 95% CI 0.46 – 1.00 and a p = 0.052, as well as was the difference between the homozygous mutated (GG) patients group and the reference category (HR = 0.614; 95% CI = 0.38 – 1.00; p = 0.051) (Fig. 44).



CYP1B1*3_4326G/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
CC	89	48	43.0	0.0675	Reference		
GC	127	57	59.4		0.682	0.46 – 1.00	0.052
GG	59	24	48.6		0.614	0.38 – 1.00	0.051

Figure 44. TTP according to the CYP1B1*3_4326G/C genotypes. In orange are highlighted the borderline statistically significant log-rank and the p-values.

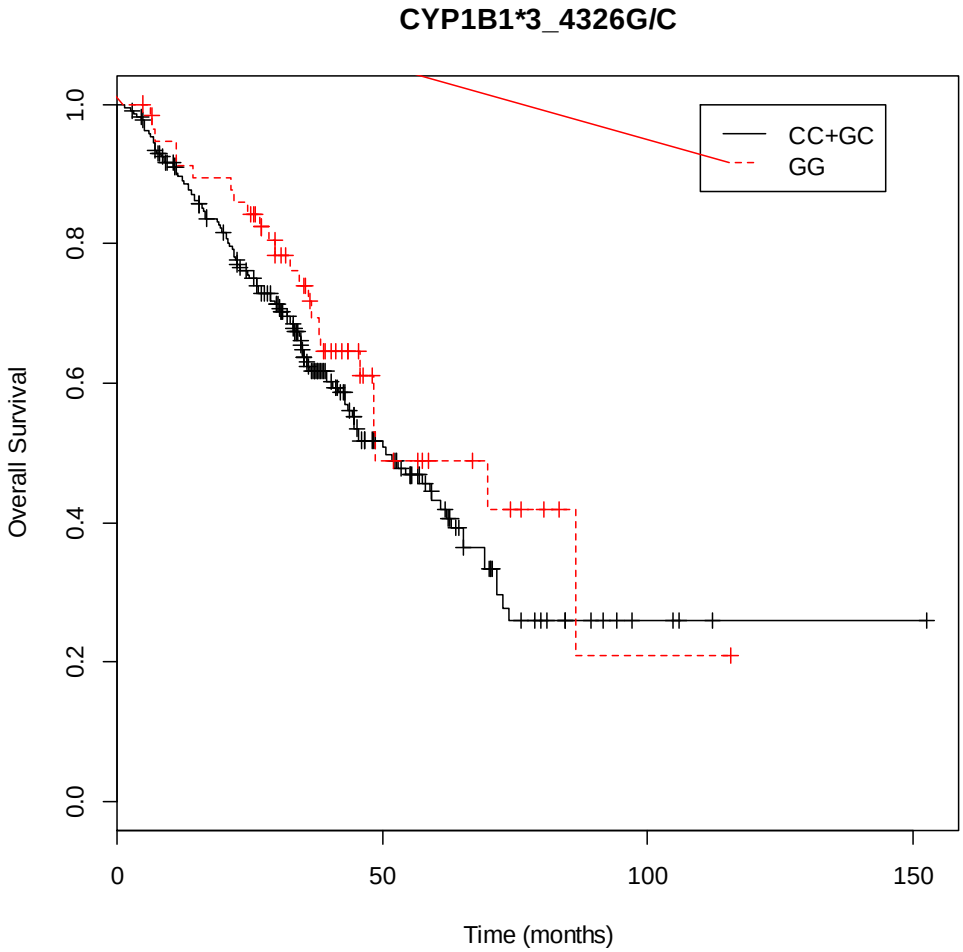
Regardless, when the dominant model was applied the result acquired full significance. Patients carrying at least one variant allele showed to have a HR of death 0.66-fold lower than the reference category group, with a median OS of 16.4 months longer (HR = 0.66; 95% CI = 0.46 – 0.95, p = 0.023) (Fig. 45).



CYP1B1*3_4326G/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
DOMINANT MODEL							
CC	89	48	43.0	0.0222	Reference		
GC+GG	186	81	59.4		0.66	0.46 – 0.95	0.023

Figure 45. OS according to the dominant model for the CYP1B1*3_4326G/C SNP. In red are highlighted the statistically significant log-rank and p-value for the comparison between the CC group and patients carrying at least one G allele.

On the contrary, with the recessive model, the median OS between the two groups of patients diverged only for 2.1 months, and this difference was not considered statistically significant (HR = 0.769; 95% CI =0.49 – 1.20; p = 0.25) (Fig. 46).

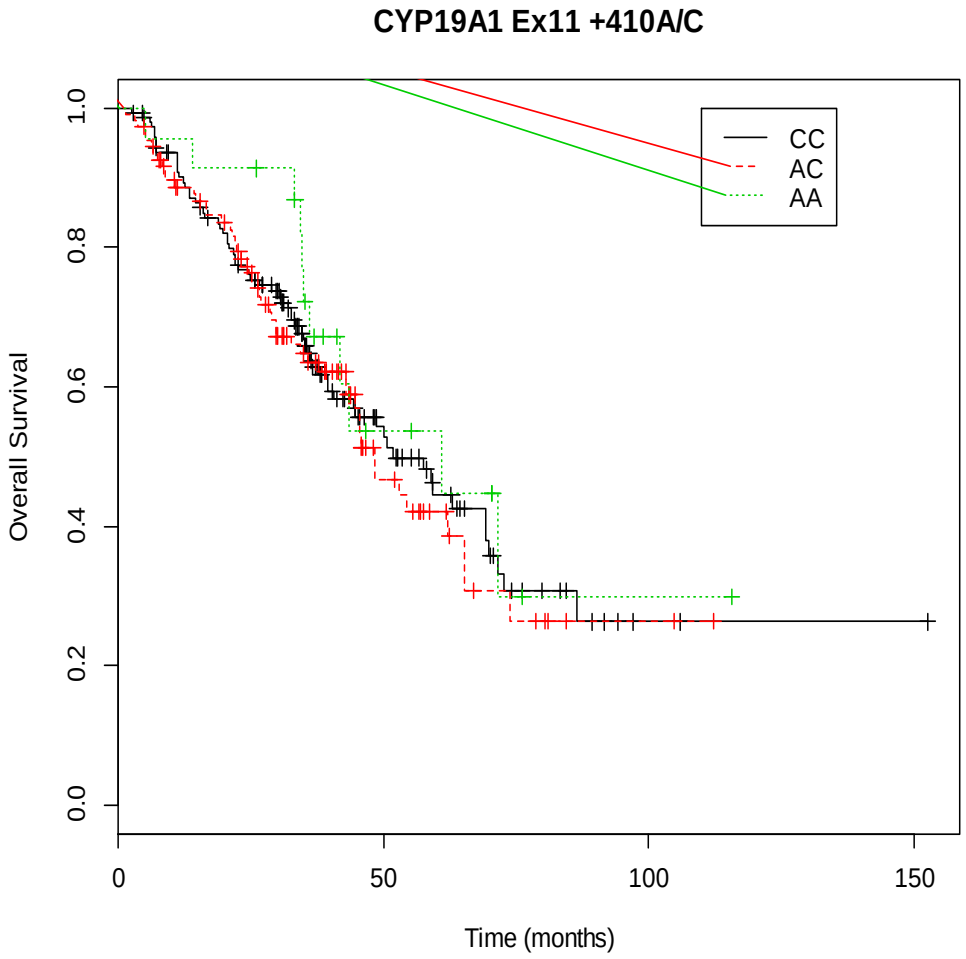


CYP1B1*3_4326G/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
RECESSIVE MODEL							
CC+GC	216	105	50.7	0.245	Reference		
GG	59	24	48.6		0.769	0.49 – 1.20	0.25

Figure 46. OS according to the recessive model for the CYP1B1*3_4326G/C SNP.

✓ **CYP19A1_Ex11_410A/C (rs4646)**

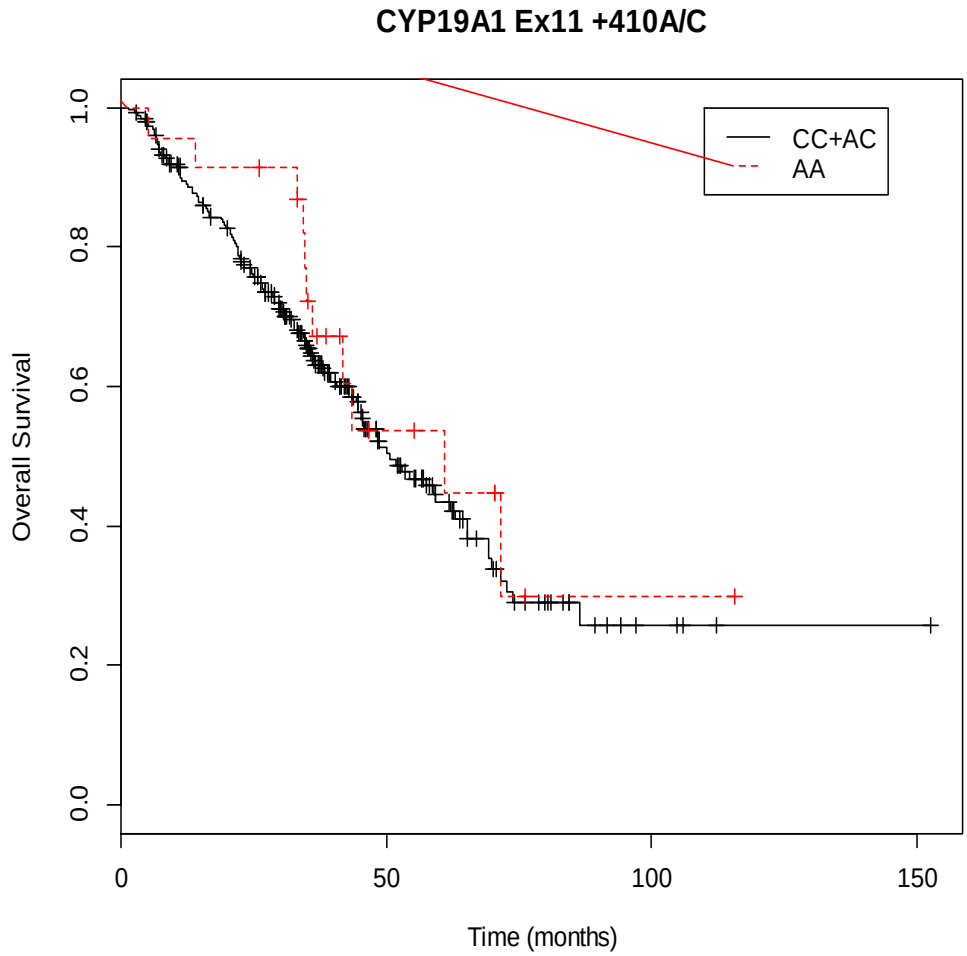
As can be seen in the Figure 47, for the CYP19A1_Ex11_410A/C (rs4646) SNP no statistically significant associations have been observed between the genotypes and the patients' OS. Death occurred in 129 cases: among them, 68 patients were homozygous for the wild type genotype (CC), 50 were heterozygous and 11 carried the mutated AA genotype. The difference of the risks of death was not statistically significant among the three groups as demonstrated by the HRs (HR_{AC} = 1.088; 95% CI of 0.75 – 1.57, p = 0.65; HR_{AA} = 0.821; 95% CI = 0.43 – 1.55, p = 0.54).



CYP19A1_EX11+410A/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
CC	144	68	51.9	0.686	Reference		
AC	108	50	48.3		1.088	0.75 -1.57	0.65
AA	23	11	60.9		0.821	0.43 -1.55	0.54

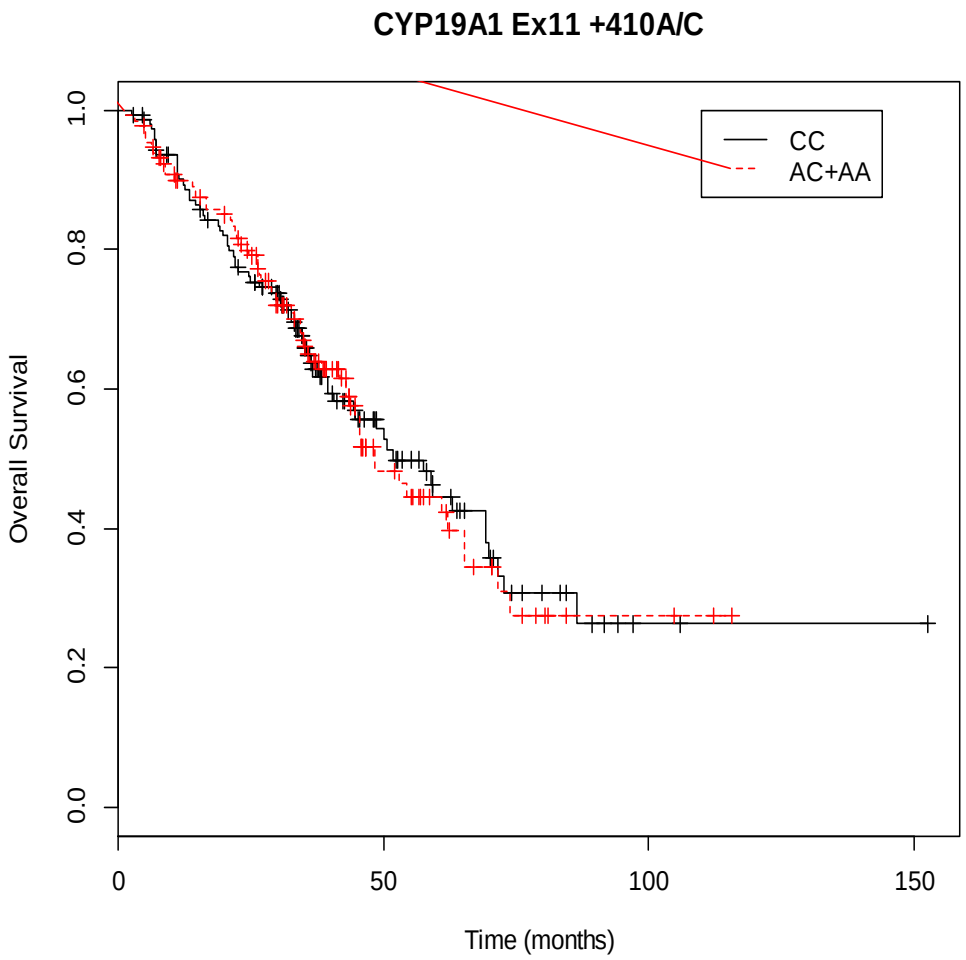
Figure 47. OS according to the CYP19A1_Ex11+410A/C genotypes.

Statistically significant associations have not been observed neither when the two genetic models (dominant and recessive) were applied (Fig. 48 and 49).



CYP19A1_EX11+410°/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
RECESSIVE MODEL							
CC+AC	252	118	50.7	0.46	Reference		
AA	23	11	60.9		0.793	0.43 – 1.47	0.46

Figure 48. OS according to the recessive model for the CYP19A1_Ex11+410A/C SNP.



CYP19A1_EX11+410°/C	N° cases	N° events	Median	p-value log-rank	HR	95%IC	p-value
DOMINANT MODEL							
CC	144	68	51.9	0.88	Reference		
AC+AA	131	61	48.3		1.03	0.73 -1.45	0.88

Figure 49. OS according to the dominant model for the CYP19A1_Ex11+410A/C SNP.

6.DISCUSSION

Personalized medicine has been defined by the US President's Council of Advisors on Science and technology as "*the tailoring of medical treatment to the individual characteristics of each patient; to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment so that preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not*". To this definition, the Personalized Medicine Coalition has added, "*the molecular methods that make personalized medicine possible include testing for variations in genes, gene expression, proteins, and metabolites, as well as new treatments that target molecular mechanisms. Test results are correlated with clinical factors – such as disease state, prediction of future disease states, drug response, and treatment prognosis – to help physicians individualize treatment for each patient*" (Personalized Medicine Coalition).

The commonly used anticancer drugs show a significant inter-individual variability in patient's outcome, in regard to both the tumor cell response and the toxic effects on non-neoplastic tissues. Such inter-subject variation is determined by several factors, which also include the individual genetic characteristics. Pharmacogenetics, analyzing the correlation between patients' genotype and phenotype, has the aim to define the relationship between a given polymorphic variation and the drug effect alteration. Accordingly, predictive or prognostic genetic biomarkers can be identified and, once validated, can be used in the clinical practice to customize and optimize the therapy by decreasing its toxicity and, at the same time, increasing its effectiveness.

Exemestane is an anti-estrogen drug that acts by inactivating aromatase, the enzyme that locally catalyzes the androgens conversion into estrogens. Estrogens stimulate tumor proliferation through the activation of receptors ER1 and ER2 on hormone-sensitive BCs. The importance of aromatase in the physiopathology of hormone-dependent tumors lies on the fact that, in postmenopausal women, this enzyme represents the only source of estrogens, since the ovarian production has been stopped. This assumption was the basis of the aromatase inhibitors development, performing these drugs their anticancer activity by suppressing the estrogens production and, therefore, their support to the ER+ cancer cells growth.

The exemestane treatment of patients affected by ER+ metastatic or locally advanced BC has proved to be a viable therapeutic strategy after the failure of a previous anti-estrogen therapy; however, in a small number of patients this drug appears to be ineffective or even toxic.

This PhD thesis had a dual aim: on one hand to set up a pharmacogenetic method to analyze estrogen- and exemestane-related SNPs and STRs, on the other hand to determine the predictive value of such polymorphisms as biomarkers of exemestane treatment efficacy. Polymorphisms in two main classes of genes were assessed: the first class involved the estrogens pathway, and, in particular, their synthesis

(*CYP19A1* and *CYP17A1*), activity (*ESR1* and *ESR2*) and metabolism (*CYP1B1*, *UGT1A1* and *COMT*); the other one was associated with the metabolic pathway of exemestane (*CYP3A4* and *CYP3A5*).

As result of the first aim of this study, a pharmacogenetic strategy, involving the application of 4 genotyping methods was developed in order to investigate the planned SNPs and STRs. In particular, according to the best results obtained in the setting up phase of the analyses, the most appropriate technique to study each polymorphisms was chosen. In particular:

- three SNPs were investigated with Pyrosequencing: CYP19A1_47T/C (rs700519), CYP3A4*1B_-392A/G (rs2740574), and RIZ1_delP704 (rs2308040);
- ten SNPs were genotyped with TaqMan® Allelic Discrimination Assay: CYP19A1_Ex11+410A/C (rs4646), CYP19A1_1558T/C (rs10046), CYP3A5*3_6986A/G (rs776746), COMT_12A/G (rs4680), ESR1_497T/C (rs2234693), ESR1_256A/G (rs9340799), ESR2_1082A/G (rs1256049), ESR2_1730A/G (rs4986938), CYP17A1_27A/G (rs743572) and CYP1B1*3_4326G/C (rs1056836);
- two STR were analyzed with Automated Fragment Analysis: CYP19A1_(TTTA)n (rs60271534) and UGT1A1*28_ TA(6/7) (rs8175347)
- twelve of the above mentioned polymorphisms were additionally analyzed by Illumina GoldenGate Assay as positive controls showing a 100% accordance within the obtained genotypes.

Concerning the second aim of this work, three SNPs, with different roles, emerged from this study:
one on *CYP1B1* gene and two on aromatase (*CYP19A1*) gene.
The drug effect was evaluated by grouping the clinical responses obtained as Response Rate (RR = complete or partial response) and as Clinical Benefit (CB = complete or partial response or stable disease) (see Materials and Methods section).
The main obtained result, according to the preliminary data collected to date, regarded a significant association between the variant CYP1B1*3_4326G/C (rs1056836) allele and the clinical response to exemestane-based anti-aromatase therapy (RR, OR_{GG} = 2.91, 95% CI = 5.88 – 1.25, p = 0.0039; according to the two-sided Fisher's exact test). The same variant allele was significantly associated with the TTP and OS too (TTP, dominant model: HR_{CG+GG}= 0.66, 95% CI = 0.50 – 0.87, p = 0.0037; OS, dominant model = HR_{CG+GG}= 0.66, 95% CI = 0.46 – 0.95, p = 0.023, according to the log-rank test).
In fact, carriers of the variant G allele (both in one or two copies), had a more beneficial drug effect than patients with wild-type CC genotype. As the number of variant alleles increased, the percentage of patients with a good response increased too. The responding heterozygous patients were the 38% compared with the 22% of the wild type carrier patients (RR: OR = 2.10, 95% CI = 3.85 – 1.14, p = 0.0179) while the percentage of responding mutated patients was 46%, more than double than the wild type

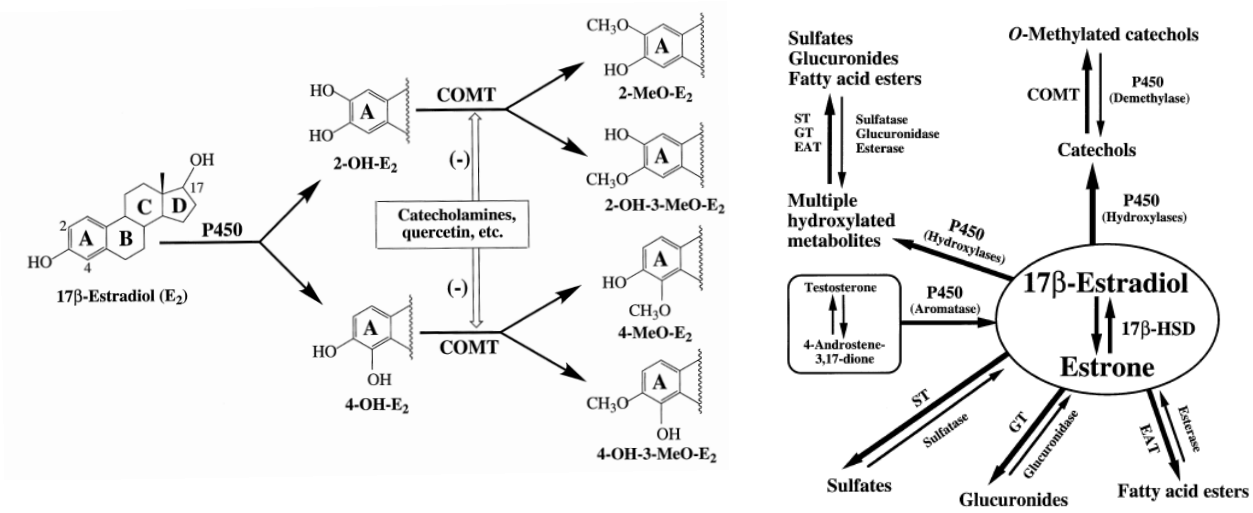
(RR: OR = 2.91, 95% CI = 5.88 – 1.25 p = 0.0039). The result was confirmed by both genetic models applied, but predominantly by the dominant one (RR, dominant model: OR = 2.33, 95% CI = 4.17 – 1.31 p = 0.0043).

The same SNP was associated with TTP and OS, too. In particular, the median TTP increased in a statistically significant manner as the number of mutated G alleles increased: indeed, heterozygous patients had a median TTP of 17.3 months compared to the 11.8 months of the wild type CC patients (HR_{CG} = 0.66, 95% CI = 0.49 – 0.89, p = 0.0073) and GG patients had a slightly longer median TTP of 18.0 months (HR_{GG} = 0.66, 95% CI = 0.45 – 0.96, p = 0.0280). Even in this case, the dominant model, already associated with RR, showed a statistically significant difference in the median TTP among patients carrying at least one G allele compared with the wild type CC patients (HR_{CG+GG} = 0.66, 95% CI = 0.50 – 0.87, p = 0.0037).

In addition, the CYP1B1*3_4326G/C (rs1056836) SNP was tendentially linked with a longer OS. Indeed, heterozygous genotype patients experienced a 59.4 months median OS compared with the 43 months observed in the CC patients group (HR_{CG}=0.68, 95% CI = 0.46 – 1.00, p = 0.052), while the mutated GG genotypes had a 48.6 months long OS (HR = 0.614, 95% CI = 0.38 – 1.00, p = 0.051). These results showed a clear tendency to significance and were confirmed with a fully significant association when the dominant model was applied (HR_{CG+GG} = 0.66, 95% CI = 0.46 – 0.95, p = 0.023). It is noteworthy that the dominant model was associated to the best clinical outcome in all three analyses (RR, TTP and OS), highlighting the consistency of the our results.

It was particularly interesting to establish the association between the variant G allele of CYP1B1*3_4326G/C (rs1056836) SNP and both TTP and OS since it highlighted a prognostic value of this SNP already associated with the clinical outcome. Overall, these results are meaningful since the same variant allele is associated with a better clinical outcome, a longer TTP and OS than wild type genotype.

It is well described in literature that CYP1B1 is differentially expressed among tissues, with the highest level in extra-hepatic ones. Moreover, CYP1B1 is present at high levels in a wide variety of tumors including those arising from hormone-responsive tissues such as BC tissue (reviewed by Gajjar *et al.*, 2012). This enzyme shows catalytic activity towards several substrates, including environmental procarcinogens and steroid hormones. CYP1B1 is currently thought to be the most efficient estrogen hydroxylase (Newbold *et al.*, 2000) responsible for local catalyzation of the estradiol 4-hydroxylation, and, in a lesser extent, 2-hydroxylation. This passage represents the first step towards the metabolic inactivation of estrogens into derivatives no longer able to stimulate the receptors. In fact, the hydroxylated estrogen metabolites, may subsequently be inactivated by O-methylation (COMT enzyme), esterification with fatty acids (Acyl-CoA), glucuronidation (UGT1A1 enzyme) or sulfonation (Zhu and Conney, 1998). If they are not eliminated by these processes, however, their sequential oxidation turns them into semiquinone and quinone derivatives, having genotoxic activity. (Gajjar *et al.*, 2012) (Fig. 40).



(a) (b)

Figure 40. Metabolic inactivation of 17β estradiol by subsequential hydroxylations and O-methylation (a) and the complexities of estrogen metabolism (b). Abbreviations: 2-OH-E_2, 2-hydroxyestradiol; 4-OH-E_2, 4-hydroxyestradiol; 2-MeO-E_2, 2-methoxyestradiol; 2-OH-3-MeO-E_2, 2-hydroxy-3-methoxyestradiol; 4-MeO-E_2, 4-methoxyestradiol; 4-OH-3-MeO-E_2, 4-hydroxy-3-methoxyestradiol, ST (sulfotransferase), GT (glucuronosyltransferase), EAT (estrogen acyltransferase; for fatty acid ester formation), 17b-HSD (17b- hydroxysteroid dehydrogenase), COMT (catechol-O-methyltransferase) and P450 (cytochrome P450).

Thus, CYP1B1 activity converts estradiol into less active catechol-estrogen derivatives (Hayes *et al.*, 1996; Shimada *et al.*, 1999; Badawi *et al*., 2001; Lee *et al.*, 2003).

It is, moreover, known that the CYP1B1*3_4326G/C (rs1056836) SNP, resulting in a (C>G) leucine to valine replacement in position 432 of the enzyme, produces a functional effect on the activity of the protein: the amino acid substitution, in fact, produces an increased catalytic activity of CYP1B1 (Landi *et al.*, 2005; Shimada *et al.*, 1999).

We speculate that the increased catalytic activity of the variant enzyme can enhance the estradiol metabolism resulting in the reduction of its local amount available to ER stimulation, thus translating into a minor tumor growth. This effect could, in turn, be reflected into a better RR, a longer TTP and OS of patients carrying the variant enzyme with respect of the wild type ones. Moreover, it is already described that estrogen metabolites can, in some way, affect the efficacy of anticancer drugs (Laroche-Clary *et al.*, 2010; Pastina *et al.*, 2010). Nevertheless, our hypothesis needs to be deepened by further studies.

The limit of our study is that we investigated the role of only one SNP on the *CYP1B1* gene, the CYP1B1*3_4326G/C (rs1056836). It has been reported that the combination with other *CYP1B1* alleles in both exons 2 and 3 may be relevant (Aklillu *et al.*, 2002), so it will be interesting to analyze other *CYP1B1* SNPs and assess their role in the clinical outcome of exemestane therapy.

Regarding the aromatase gene, besides the CYP19A1_Ex11_410A/C (rs4646) SNP, the main objective of this study, three other SNPs were investigated: CYP19A1_1558T/C (rs10046), CYP19A1_47T/C (rs700519), CYP19A1_(TTTA)n (rs60271534). The only association found in our study regarded the

CYP19A1_1558T/C (rs10046) SNP, which resulted to be linked, although in a marginal manner, to the TTP (HR_{CC} recessive model=1.4, 95%CI = 1.04 – 1.89, p = 0.028).

The role of aromatase gene polymorphisms is controversial, since many and discordant results have been reported in literature for their association with the clinical outcome of BC patients.

To the best of our knowledge, among several studies which have assessed the role of CYP19A1_1558T/C (rs10046) SNP, only one showed an association with BC patients disease free survival (DFS). In that study, carried out in 1257 invasive primary BC patients, the mutated CC genotype was associated with a better prognosis than the TT one (Fasching *et al.*, 2008). On the contrary, in our study, the mutated CC genotype was tendentially associated with a shorter median DFS (assessed as TTP) (HR_{CC}=1.41, 95% CI = 0.99 – 2.00, p = 0.056) and the result was fully significant when the recessive model was applied (HR=1.4, 95% CI = 1.04 – 1.89, p = 0.028).

The effect observed in the Fasching's study appeared to be linked to the menopausal status, since it derived only from the premenopausal patients' subgroup. This could be the reason for its divergence from our study, which was carried out on postmenopausal patients only. It could be possible that, depending on the patient's menopausal status, the C allele exerts an opposite effect on the prognosis. Another aspect to consider is that the treatment received by patients was not specified in Fasching's work. This could be another source for the discrepancy with our study in which the patients were homogeneously treated with exemestane.

This polymorphism, lying on the 3'UTR, has been associated with the mRNA level of CYP19A1 enzyme: in particular a reduced mRNA level has been linked to the C allele (Kristensen *et al.*, 2000). Nevertheless, despite this functional effect, we hypothesize that the estrogenic synthesis inhibition produced by exemestane in postmenopausal patients (which, for their postmenopausal condition have already a lower baseline estradiol concentration than premenopausal women) could be strong enough to override any possible variation exercised by aromatase SNPs on estrogen synthesis.

Another interesting finding of our study was the observation of a new genetic variant for the CYP19A1_(TTTA)n (rs60271534) STR in the intron 4 of the aromatase gene.

The number of repeats has been associated to many clinical conditions, for example, to BC risk (Bexter *et al.*, 2001, Han *et al.*, 2005, Ribeiro *et al.*, 2006), prognosis of BC patients (Huang *et al.*, 2008, Kuo *et al.*, 2013) and bone mineral density (Markatseli *et al.*, 2014). Our study did not show any association of this polymorphism with patients' clinical outcome.

Genetic databases and literature report that the number of the repeats varies from 7 to 13, whereas we detected a still not described 14 (TTTA) repeats allele. Another study, conducted in Brazilian people described a patient with a 6 repeats allele, meaning that there is still room for more investigations (Ribiero *et al.*, 2006).

The primary objective of this study was to assess the predictive role of CYP19A1_Ex11_410A/C (rs4646) SNP in the aromatase gene (*CYP19A1*) on response to treatment with exemestane.

Statistical analysis showed no significant correlation between this SNP and the obtained response, neither in terms of RR nor CB. There was also a lack of association between the SNP and TTP or OS. The role of this polymorphism has not yet been clearly defined, since it has been variously associated to the clinical outcome of several AIs used in different therapeutic settings and with highly discordant results.

A positive association between the variant A allele and the clinical response was reported by Colomer and collaborators in the work taken as reference at the time of starting of this study. In that work, performed on 67 patients treated with the AI letrozole, the variant genotype of CYP19A1_Ex11_410A/C SNP was associated with a longer PFS (Colomer *et al.*, 2008). In particular, patients with at least one variant allele (AC/AA) showed a three times longer TTP than patients homozygous for the reference allele (CC). In addition, the authors reported that the frequency of the variant A allele was significantly higher in responding patients (61%) than in the non-responders group (40%). The predictive value of the AA genotype has also been shown for another AI, anastrozole. In the study by Liu and colleagues (Liu *et al.*, 2013), the variant A allele was associated with improved OS (37.3 months versus 31.6 months; p = 0.007) in 272 MBC patients treated with anastrozole in first metastatic line after progression following tamoxifen.

On the contrary, the same variant was associated with a lower therapeutic benefit in terms of DFS in a cohort of 95 patients treated with neoadjuvant letrozole (Garcia-Casado *et al.*, 2010).

Finally, no significant association regarding the polymorphism and the clinical outcome was observed by Ferraldeschi and collaborators (Ferraldeschi *et al.*, 2012) on a population of 308 patients affected by metastatic BC treated with a third-generation aromatase inhibitor.

Considering the heterogeneity of data reported in literature, the interpretation of our results becomes complex. The first aspect to consider is that there is no uniformity in the source of DNA used for the analysis and this could explain the reported differences between our results and those obtained by Colomer. In fact, in our study we used germ-line DNA extracted from peripheral blood, while Colomer's group conducted genetic analyses on DNA extracted from paraffin-embedded tumor tissue. Cancer tissues undergo genetic alterations that may also involve the gene portion under study. The CYP19A1 is localized in the region 21.2 of the long arm of chromosome 15 (15q21.2) and has been stated that this area is a frequent target of allelic imbalance in advanced BC, that can affect the distribution of allelic variants (Garcia-Casado *et al.*, 2010). Regarding the result reported by Garcia-Casado, the therapy setting, neoadjuvant instead of the first metastatic line, may account for the different obtained results. In particular, the analyzed SNP could have a prognostic and predictive significance only in a given therapeutic setting.

Overall, we hypothesize that the different effect on the treatment outcome could be related to the local estradiol concentration. As above mentioned, we speculate that the estrogenic synthesis inhibition produced by exemestane could be strong enough to override any possible variation exercised by aromatase SNPs on estrogen synthesis. On the other side, in this context, a more active estradiol metabolism, exerted by the CYP1B1 variant enzyme, could impact the local estradiol amount and, thus, affect patients' clinical outcome, although this hypothesis needs to be confirmed by further analyses.

It remains to be verified the role of the so far investigated SNPs on affecting the toxicity of exemestane treatment and the impact of the GoldenGate assay SNPs in both exemestane efficacy (RR, TTP and OS) and safety.

7.CONCLUSIONS

In conclusion, the first aim of this PhD thesis was to develop a pharmacogenetic strategy to investigate 15 exemestane-related SNPs. We set up a series of 4 methodologies employed to analyze the SNPs and cross-validate the obtained results. This strategy was fully successful and allowed the genotyping and the validation of the SNPs investigated.

On the other hand, the aim of this study was to determine whether polymorphisms involved in exemestane and estrogens activity and metabolism could be used as predictive markers of response to treatment. Our study showed that the CYP1B1*3_4326G/C (rs1056836) was predictive of the clinical outcome of metastatic or locally advanced BC patients treated with exemestane as a first line hormone therapy. This SNP had also a prognostic value as regards the survival. Indeed, patients carrying at least one variant allele (G) not only showed a better clinical response, but experienced also a later progression and a longer survival than wild type patients. This indicates that a simple genetic evaluation from peripheral blood, performed prior to therapy, may allow the identification of patients more likely to be responsive to treatment.

So, this study led to the identification of a predictive genetic marker of exemestane response, which, once validated, could potentially be employable in the daily clinical oncology practice.

Future perspectives will be the characterization of the other SNPs on the *CYP1B1* gene and their association with patients' clinical outcome, the verification of the role of the so far investigated SNPs on affecting exemestane toxicity and the evaluation of the impact of the GoldenGate assay SNPs, in both exemestane efficacy (RR, TTP and OS) and safety.

8. REFERENCES

Adegoke OJ, Shu XO, Gao YT, Cai Q, Breyer J, Smith J, et al. Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 (UGT1A1) and risk of breast cancer. Breast Cancer Res Treat 2004 Jun;85(3):239-45.

Aklillu E, Oscarson M, Hidestrand M, Leidvik B, Otter C, Ingelman-Sundberg M. Functional analysis of six different polymorphic CYP1B1 enzyme variants found in an Ethiopian population. Mol Pharmacol 2002 Mar;61(3):586-94.

Albert C, Vallee M, Beaudry G, Belanger A, Hum DW. The monkey and human uridine diphosphate-glucuronosyltransferase UGT1A9, expressed in steroid target tissues, are estrogen-conjugating enzymes. Endocrinology 1999 Jul;140(7):3292-302.

Badawi AF, Cavalieri EL, Rogan EG. Role of human cytochrome P450 1A1, 1A2, 1B1, and 3A4 in the 2-, 4-, and 16alpha-hydroxylation of 17beta-estradiol. Metabolism 2001 Sep;50(9):1001-3.

Baxter SW, Choong DY, Eccles DM, Campbell IG. Polymorphic variation in CYP19 and the risk of breast cancer. Carcinogenesis 2001 Feb;22(2):347-9.

Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? Proc Natl Acad Sci U S A 1998 Jul 7;95(14):8170-4.

BLOOM HJ, RICHARDSON WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957 Sep;11(3):359-77.

Borgquist S, Hjertberg M, Henningson M, Ingvar C, Rose C, Jernstrom H. Given breast cancer, is fat better than thin? Impact of the estrogen receptor beta gene polymorphisms. Breast Cancer Res Treat 2013 Feb;137(3):849-62.

Bosma PJ, Chowdhury JR, Bakker C, Gantla S, de BA, Oostra BA, et al. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med 1995 Nov 2;333(18):1171-5.

Bosma PJ. Inherited disorders of bilirubin metabolism. J Hepatol 2003 Jan;38(1):107-17.

Brueggemeier RW, Hackett JC, Diaz-Cruz ES. Aromatase inhibitors in the treatment of breast cancer. Endocr Rev 2005 May;26(3):331-45.

Brueggemeier RW. Update on the use of aromatase inhibitors in breast cancer. Expert Opin Pharmacother 2006 Oct;7(14):1919-30.

Butler JM. Short tandem repeat analysis for human identity testing. Curr Protoc Hum Genet 2004 Sep;Chapter 14:Unit.

Butler JM, Buel E, Crivellente F, McCord BR. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 2004 Jun;25(10-11):1397-412.

Buzdar AU, Robertson JF, Eiermann W, Nabholtz JM. An overview of the pharmacology and pharmacokinetics of the newer generation aromatase inhibitors anastrozole, letrozole, and exemestane. Cancer 2002 Nov 1;95(9):2006-16.

Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2012 Oct;23 Suppl 7:vii11-vii19.

Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006 Jun 7;295(21):2492-502.

Chia S, Gradishar W, Mauriac L, Bines J, Amant F, Federico M, et al. Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: results from EFECT. J Clin Oncol 2008 Apr 1;26(10):1664-70.

Chia SK, Speers CH, D'yachkova Y, Kang A, Malfair-Taylor S, Barnett J, et al. The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. Cancer 2007 Sep 1;110(5):973-9.

Coleman RE, Banks LM, Girgis SI, Kilburn LS, Vrdoljak E, Fox J, et al. Skeletal effects of exemestane on bone-mineral density, bone biomarkers, and fracture incidence in postmenopausal women with early breast cancer participating in the Intergroup Exemestane Study (IES): a randomised controlled study. Lancet Oncol 2007 Feb;8(2):119-27.

Colomer R, Monzo M, Tusquets I, Rifa J, Baena JM, Barnadas A, et al. A single-nucleotide polymorphism in the aromatase gene is associated with the efficacy of the aromatase inhibitor letrozole in advanced breast carcinoma. Clin Cancer Res 2008 Feb 1;14(3):811-6.

Coombes RC, Hall E, Gibson LJ, Paridaens R, Jassem J, Delozier T, et al. A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. N Engl J Med 2004 Mar 11;350(11):1081-92.

Coombes RC, Kilburn LS, Snowdon CF, Paridaens R, Coleman RE, Jones SE, et al. Survival and safety of exemestane versus tamoxifen after 2-3 years' tamoxifen treatment (Intergroup Exemestane Study): a randomised controlled trial. Lancet 2007 Feb 17;369(9561):559-70.

Cruz JJ, Richart AP, Garcia MJ, Fernandez MR, Pelaez F, I, Sampedro GT, et al. Management of patients with metastatic breast cancer. Adv Ther 2011 Sep;28 Suppl 6:50-65.

Cruz JJ, Richart AP, Garcia MJ, Fernandez MR, Pelaez F, I, Sampedro GT, et al. Management of patients with metastatic breast cancer. Adv Ther 2011 Sep;28 Suppl 6:50-65.

Dai D, Tang J, Rose R, Hodgson E, Bienstock RJ, Mohrenweiser HW, et al. Identification of variants of CYP3A4 and characterization of their abilities to metabolize testosterone and chlorpyrifos. J Pharmacol Exp Ther 2001 Dec;299(3):825-31.

Dawling S, Roodi N, Mernaugh RL, Wang X, Parl FF. Catechol-O-methyltransferase (COMT)-mediated metabolism of catechol estrogens: comparison of wild-type and variant COMT isoforms. Cancer Res 2001 Sep 15;61(18):6716-22.

Dawood S, Broglio K, Buzdar AU, Hortobagyi GN, Giordano SH. Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. J Clin Oncol 2010 Jan 1;28(1):92-8.

Deenen MJ, Cats A, Beijnen JH, Schellens JH. Part 1: background, methodology, and clinical adoption of pharmacogenetics. Oncologist 2011;16(6):811-9.

Dick IM, Devine A, Prince RL. Association of an aromatase TTTA repeat polymorphism with circulating estrogen, bone structure, and biochemistry in older women. Am J Physiol Endocrinol Metab 2005 May;288(5):E989-E995.

Efferth T, Volm M. Pharmacogenetics for individualized cancer chemotherapy. Pharmacol Ther 2005 Aug;107(2):155-76.

Ekhart C, Rodenhuis S, Smits PH, Beijnen JH, Huitema AD. An overview of the relations between polymorphisms in drug metabolising enzymes and drug transporters and survival after cancer drug treatment. Cancer Treat Rev 2009 Feb;35(1):18-31.

Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. C. W. Elston & I. O. Ellis. Histopathology 1991; 19; 403-410. Histopathology 2002 Sep;41(3A):151-2, discussion.

Fakhrai-Rad H, Pourmand N, Ronaghi M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. Hum Mutat 2002 May;19(5):479-85.

Fasching PA, Loehberg CR, Strissel PL, Lux MP, Bani MR, Schrauder M, et al. Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients. Breast Cancer Res Treat 2008 Nov;112(1):89-98.

Ferraldeschi R, Arnedos M, Hadfield KD, A'Hern R, Drury S, Wardley A, et al. Polymorphisms of CYP19A1 and response to aromatase inhibitors in metastatic breast cancer patients. Breast Cancer Res Treat 2012 Jun;133(3):1191-8.

Fu C, Dong WQ, Wang A, Qiu G. The influence of ESR1 rs9340799 and ESR2 rs1256049 polymorphisms on prostate cancer risk. Tumour Biol 2014 Aug;35(8):8319-28.

Gajjar K, Martin-Hirsch PL, Martin FL. CYP1B1 and hormone-induced cancer. Cancer Lett 2012 Nov 1;324(1):13-30.

Garcia-Casado Z, Guerrero-Zotano A, Llombart-Cussac A, Calatrava A, Fernandez-Serra A, Ruiz-Simon A, 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.

Genestie C, Zafrani B, Asselain B, Fourquet A, Rozan S, Validire P, et al. Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. Anticancer Res 1998 Jan;18(1B):571-6.

Gennari A, Conte P, Rosso R, Orlandini C, Bruzzi P. Survival of metastatic breast carcinoma patients over a 20-year period: a retrospective analysis based on individual patient data from six consecutive studies. Cancer 2005 Oct 15;104(8):1742-50.

Geyer FC, Marchio C, Reis-Filho JS. The role of molecular analysis in breast cancer. Pathology 2009 Jan;41(1):77-88.

Giordano SH, Buzdar AU, Smith TL, Kau SW, Yang Y, Hortobagyi GN. Is breast cancer survival improving? Cancer 2004 Jan 1;100(1):44-52.

Goss PE, Qi S, Cheung AM, Hu H, Mendes M, Pritzker KP. Effects of the steroidal aromatase inhibitor exemestane and the nonsteroidal aromatase inhibitor letrozole on bone and lipid metabolism in ovariectomized rats. Clin Cancer Res 2004 Sep 1;10(17):5717-23.

Greenberg PA, Hortobagyi GN, Smith TL, Ziegler LD, Frye DK, Buzdar AU. Long-term follow-up of patients with complete remission following combination chemotherapy for metastatic breast cancer. J Clin Oncol 1996 Aug;14(8):2197-205.

Grundberg E, Carling T, Brandstrom H, Huang S, Ribom EL, Ljunggren O, et al. A deletion polymorphism in the RIZ gene, a female sex steroid hormone receptor coactivator, exhibits decreased response to estrogen in vitro and associates with low bone mineral density in young Swedish women. J Clin Endocrinol Metab 2004 Dec;89(12):6173-8.

Guarneri V, Conte P. Metastatic breast cancer: therapeutic options according to molecular subtypes and prior adjuvant therapy. Oncologist 2009 Jul;14(7):645-56.

Guillemette C, Millikan RC, Newman B, Housman DE. Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and association with breast cancer among African Americans. Cancer Res 2000 Feb 15;60(4):950-6.

Han DF, Zhou X, Hu MB, Xie W, Mao ZF, Chen DE, et al. Polymorphisms of estrogen-metabolizing genes and breast cancer risk: a multigenic study. Chin Med J (Engl) 2005 Sep 20;118(18):1507-16.

Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, et al. Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse. Breast Cancer Res Treat 2012 Apr;132(2):523-35.

Hayes CL, Spink DC, Spink BC, Cao JQ, Walker NJ, Sutter TR. 17 beta-estradiol hydroxylation catalyzed by human cytochrome P450 1B1. Proc Natl Acad Sci U S A 1996 Sep 3;93(18):9776-81.

Hoggatt J. Personalized medicine--trends in molecular diagnostics: exponential growth expected in the next ten years. Mol Diagn Ther 2011 Feb 1;15(1):53-5.

Hoskins JM, Goldberg RM, Qu P, Ibrahim JG, McLeod HL. UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. J Natl Cancer Inst 2007 Sep 5;99(17):1290-5.

Huang CS, Kuo SH, Lien HC, Yang SY, You SL, Shen CY, et al. The CYP19 TTTA repeat polymorphism is related to the prognosis of premenopausal stage I-II and operable stage III breast cancers. Oncologist 2008 Jul;13(7):751-60.

Innocenti F, Ratain MJ. Irinotecan treatment in cancer patients with UGT1A1 polymorphisms. Oncology (Williston Park) 2003 May;17(5 Suppl 5):52-5.

Ioannidis JP, Ralston SH, Bennett ST, Brandi ML, Grinberg D, Karassa FB, et al. Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. JAMA 2004 Nov 3;292(17):2105-14.

Iolascon A, Faienza MF, Centra M, Storelli S, Zelante L, Savoia A. (TA)8 allele in the UGT1A1 gene promoter of a Caucasian with Gilbert's syndrome. Haematologica 1999 Feb;84(2):106-9.

Johnston SR, Kilburn LS, Ellis P, Dodwell D, Cameron D, Hayward L, et al. Fulvestrant plus anastrozole or placebo versus exemestane alone after progression on non-steroidal aromatase inhibitors in postmenopausal patients with hormone-receptor-positive locally advanced or metastatic breast cancer (SoFEA): a composite, multicentre, phase 3 randomised trial. Lancet Oncol 2013 Sep;14(10):989-98.

Kagimoto M, Heim M, Kagimoto K, Zeugin T, Meyer UA. Multiple mutations of the human cytochrome P450IID6 gene (CYP2D6) in poor metabolizers of debrisoquine. Study of the functional significance of individual mutations by expression of chimeric genes. J Biol Chem 1990 Oct 5;265(28):17209-14.

Kooloos WM, Wessels JA, van der Straaten T, Huizinga TW, Guchelaar HJ. Criteria for the selection of single nucleotide polymorphisms in pathway pharmacogenetics: TNF inhibitors as a case study. Drug Discov Today 2009 Sep;14(17-18):837-44.

Kristensen VN, Harada N, Yoshimura N, Haraldsen E, Lonning PE, Erikstein B, et al. Genetic variants of CYP19 (aromatase) and breast cancer risk. Oncogene 2000 Mar 2;19(10):1329-33.

Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat Genet 2001 Apr;27(4):383-91.

Kuo SH, Yang SY, Lien HC, Lo C, Lin CH, Lu YS, et al. CYP19 genetic polymorphism haplotype AASA is associated with a poor prognosis in premenopausal women with lymph node-negative, hormone receptor-positive breast cancer. Biomed Res Int 2013;2013:562197.

Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics 1996 Jun;6(3):243-50.

Lakhani S, Ellis I, Schnitt S, et al. 4th. Lyon: IARC Press; 2012. WHO Classification of Tumours of the Breast (4^{th} edition).

Landi MT, Bergen AW, Baccarelli A, Patterson DG, Jr., Grassman J, Ter-Minassian M, et al. CYP1A1 and CYP1B1 genotypes, haplotypes, and TCDD-induced gene expression in subjects from Seveso, Italy. Toxicology 2005 Feb 14;207(2):191-202.

Lange CA, Sartorius CA, Abdel-Hafiz H, Spillman MA, Horwitz KB, Jacobsen BM. Progesterone receptor action: translating studies in breast cancer models to clinical insights. Adv Exp Med Biol 2008;630:94-111.

Laroche-Clary A, Le M, V, Yamori T, Robert J. Cytochrome P450 1B1 gene polymorphisms as predictors of anticancer drug activity: studies with in vitro models. Mol Cancer Ther 2010 Dec;9(12):3315-21.

Lazarus P, Sun D. Potential role of UGT pharmacogenetics in cancer treatment and prevention: focus on tamoxifen and aromatase inhibitors. Drug Metab Rev 2010 Feb;42(1):182-94.

Lee AJ, Cai MX, Thomas PE, Conney AH, Zhu BT. Characterization of the oxidative metabolites of 17beta-estradiol and estrone formed by 15 selectively expressed human cytochrome p450 isoforms. Endocrinology 2003 Aug;144(8):3382-98.

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011 Jul;121(7):2750-67.

Liu L, Bai YX, Zhou JH, Sun XW, Sui H, Zhang WJ, et al. A polymorphism at the 3'-UTR region of the aromatase gene is associated with the efficacy of the aromatase inhibitor, anastrozole, in metastatic breast carcinoma. Int J Mol Sci 2013;14(9):18973-88.

Long JR, Kataoka N, Shu XO, Wen W, Gao YT, Cai Q, et al. Genetic polymorphisms of the CYP19A1 gene and breast cancer survival. Cancer Epidemiol Biomarkers Prev 2006 Nov;15(11):2115-22.

Lonning PE, Bajetta E, Murray R, Tubiana-Hulin M, Eisenberg PD, Mickiewicz E, et al. Activity of exemestane in metastatic breast cancer after failure of nonsteroidal aromatase inhibitors: a phase II trial. J Clin Oncol 2000 Jun;18(11):2234-44.

Lonning PE. Aromatase inhibitors in breast cancer. Endocr Relat Cancer 2004 Jun;11(2):179-89.

Lonning PE. Aromatase inhibitors in breast cancer. Endocr Relat Cancer 2004 Jun;11(2):179-89.

Lonning PE, Geisler J, Krag LE, Erikstein B, Bremnes Y, Hagen AI, et al. Effects of exemestane administered for 2 years versus placebo on bone mineral density, bone biomarkers, and plasma lipids in patients with surgically resected early breast cancer. J Clin Oncol 2005 Aug 1;23(22):5126-37.

Lonning PE, Eikesdal HP. Aromatase inhibition 2013: clinical state of the art and questions that remain to be solved. Endocr Relat Cancer 2013 Aug;20(4):R183-R201.

Lund MJ, Butler EN, Hair BY, Ward KC, Andrews JH, Oprea-Ilies G, et al. Age/race differences in HER2 testing and in incidence rates for breast cancer triple subtypes: a population-based study and first report. Cancer 2010 Jun 1;116(11):2549-59.

Markatseli AE, Lazaros L, Markoula S, Kostoulas H, Sakaloglou P, Tigas S, et al. Association of the (TTTA)n repeat polymorphism of CYP19 gene with bone mineral density in Greek peri- and postmenopausal women. Clin Endocrinol (Oxf) 2014 Jul;81(1):38-44.

Marsh S, McLeod HL. Pharmacogenomics: from bedside to clinical practice. Hum Mol Genet 2006 Apr 15;15 Spec No 1:R89-R93.

Martucci CP, Fishman J. P450 enzymes of estrogen metabolism. Pharmacol Ther 1993 Feb;57(2-3):237-57.

Meech R, Mackenzie PI. Structure and function of uridine diphosphate glucuronosyltransferases. Clin Exp Pharmacol Physiol 1997 Dec;24(12):907-15.

Mendoza N, Quereda F, Presa J, Salamanca A, Sanchez-Borrego R, Vazquez F, et al. Estrogen-related genes and postmenopausal osteoporosis risk. Climacteric 2012 Dec;15(6):587-93.

Metzger-Filho O, Sun Z, Viale G, Price KN, Crivellari D, Snyder RD, et al. Patterns of Recurrence and outcome according to breast cancer subtypes in lymph node-negative disease: results from international breast cancer study group trials VIII and IX. J Clin Oncol 2013 Sep 1;31(25):3083-90.

Monaghan G, McLellan A, McGeehan A, Li VS, Mollica F, Salemi I, et al. Gilbert's syndrome is a contributory factor in prolonged unconjugated hyperbilirubinemia of the newborn. J Pediatr 1999 Apr;134(4):441-6.

Muller PY, Milton MN. The determination and interpretation of the therapeutic index in drug development. Nat Rev Drug Discov 2012 Oct;11(10):751-61.

Newbold RR, Liehr JG. Induction of uterine adenocarcinoma in CD-1 mice by catechol estrogens. Cancer Res 2000 Jan 15;60(2):235-7.

Olson SH, Bandera EV, Orlow I. Variants in estrogen biosynthesis genes, sex steroid hormone levels, and endometrial cancer: a HuGE review. Am J Epidemiol 2007 Feb 1;165(3):235-45.

Pagani O, Senkus E, Wood W, Colleoni M, Cufer T, Kyriakides S, et al. International guidelines for management of metastatic breast cancer: can metastatic breast cancer be cured? J Natl Cancer Inst 2010 Apr 7;102(7):456-63.

Paridaens R, Dirix L, Lohrisch C, Beex L, Nooij M, Cameron D, et al. Mature results of a randomized phase II multicenter study of exemestane versus tamoxifen as first-line hormone therapy for postmenopausal women with metastatic breast cancer. Ann Oncol 2003 Sep;14(9):1391-8.

Paridaens RJ, Dirix LY, Beex LV, Nooij M, Cameron DA, Cufer T, et al. Phase III study comparing exemestane with tamoxifen as first-line hormonal treatment of metastatic breast cancer in postmenopausal women: the European Organisation for Research and Treatment of Cancer Breast Cancer Cooperative Group. J Clin Oncol 2008 Oct 20;26(30):4883-90.

Pastina I, Giovannetti E, Chioni A, Sissung TM, Crea F, Orlandini C, et al. Cytochrome 450 1B1 (CYP1B1) polymorphisms associated with response to docetaxel in Castration-Resistant Prostate Cancer (CRPC) patients. BMC Cancer 2010;10:511.

Perou CM. Molecular stratification of triple-negative breast cancers. Oncologist 2011;16 Suppl 1:61-70.

Personalized Medicine Coalition The case for personalized medicine. Personalized Medicine Coalition, Washington, DC. 2006.

Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol 2011 Feb;5(1):5-23.

Putnik M, Zhao C, Gustafsson JA, Dahlman-Wright K. Effects of two common polymorphisms in the 3' untranslated regions of estrogen receptor beta on mRNA stability and translatability. BMC Genet 2009;10:55.

Raftogianis R, Creveling C, Weinshilboum R, and Weisz J; Chapter 6: Estrogen Metabolism by Conjugation; J Natl Cancer Inst Monogr (2000) 2000 (27): 113-124

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med 2013 Sep;15(9):733-47.

Ribeiro FS, de Amorim LM, de Almeida ST, Mendonca GA, de Moura Gallo CV, Pinto LF. CYP19 (TTTA)n polymorphism and breast cancer risk in Brazilian women. Toxicol Lett 2006 Jun 20;164(1):90-5.

Robert J, Le M, V, Giovannetti E, Peters GJ. On the use of pharmacogenetics in cancer treatment and clinical trials. Eur J Cancer 2014 Oct;50(15):2532-43.

Robert J, Le M, V, Giovannetti E, Peters GJ. On the use of pharmacogenetics in cancer treatment and clinical trials. Eur J Cancer 2014 Oct;50(15):2532-43.

Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Jr., Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005 Oct 20;353(16):1673-84.

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001 Jan;11(1):3-11.

Sachse C, Brockmoller J, Bauer S, Roots I. Cytochrome P450 2D6 variants in a Caucasian population: allele frequencies and phenotypic consequences. Am J Hum Genet 1997 Feb;60(2):284-95.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977 Dec;74(12):5463-7.

Scariano JK, Simplicio SG, Montoya GD, Garry PJ, Baumgartner RN. Estrogen receptor beta dinucleotide (CA) repeat polymorphism is significantly associated with bone mineral density in postmenopausal women. Calcif Tissue Int 2004 Jun;74(6):501-8.

Shimada T, Watanabe J, Kawajiri K, Sutter TR, Guengerich FP, Gillam EM, et al. Catalytic properties of polymorphic human cytochrome P450 1B1 variants. Carcinogenesis 1999 Aug;20(8):1607-13.

Somner J, McLellan S, Cheung J, Mak YT, Frost ML, Knapp KM, et al. Polymorphisms in the P450 c17 (17-hydroxylase/17,20-Lyase) and P450 c19 (aromatase) genes: association with serum sex steroid concentrations and bone mineral density in postmenopausal women. J Clin Endocrinol Metab 2004 Jan;89(1):344-51.

Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med 2009 Feb 19;360(8):790-800.

Stanczyk FZ, Bretsky P (2003) Biosynthesis, Transport and Metabolism of Steroid Hormones; II, 12:37. Henderson BE, Ponder B and Ross RK, IN: Hormones, Genes, and Cancer. Oxford University Press, Inc, New York

Sun D, Chen G, Dellinger RW, Sharma AK, Lazarus P. Characterization of 17-dihydroexemestane glucuronidation: potential role of the UGT2B17 deletion in exemestane pharmacogenetics. Pharmacogenet Genomics 2010 Oct;20(10):575-85.

Swain SM, Kim SB, Cortes J, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. Lancet Oncol 2013 May;14(6):461-71.

Teng Y, He C, Zuo X, Li X. Catechol-O-methyltransferase and cytochrome P-450 1B1 polymorphisms and endometrial cancer risk: a meta-analysis. Int J Gynecol Cancer 2013 Mar;23(3):422-30.

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000 Feb 2;92(3):205-16.

Tofteng CL, Kindmark A, Brandstrom H, Abrahamsen B, Petersen S, Stiger F, et al. Polymorphisms in the CYP19 and AR genes--relation to bone mass and longitudinal bone changes in postmenopausal women with or without hormone replacement therapy: The Danish Osteoporosis Prevention Study. Calcif Tissue Int 2004 Jan;74(1):25-34.

Tukey RH, Strassburg CP, Mackenzie PI. Pharmacogenomics of human UDP-glucuronosyltransferases and irinotecan toxicity. Mol Pharmacol 2002 Sep;62(3):446-50.

Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010 Apr 1;28(10):1684-91.

Wan GX, Cao YW, Li WQ, Li YC, Li F. The Catechol-O-Methyltransferase Val158Met Polymorphism Contributes to the Risk of Breast Cancer in the Chinese Population: An Updated Meta-Analysis. J Breast Cancer 2014 Jun;17(2):149-56.

Xie HG, Wood AJ, Kim RB, Stein CM, Wilkinson GR. Genetic variability in CYP3A5 and its possible consequences. Pharmacogenomics 2004 Apr;5(3):243-72.

Yilmaz MB, Pazarbasi A, Guzel AI, Kocaturk-Sel S, Kasap H, Kasap M, et al. Association of serum sex steroid levels and bone mineral density with CYP17 and CYP19 gene polymorphisms in postmenopausal women in Turkey. Genet Mol Res 2011;10(3):1999-2008.

Zhang D, Zhang D, Cui D, Gambardella J, Ma L, Barros A, et al. Characterization of the UDP glucuronosyltransferase activity of human liver microsomes genotyped for the UGT1A1*28 polymorphism. Drug Metab Dispos 2007 Dec;35(12):2270-80.

Zheng W, Xie DW, Jin F, Cheng JR, Dai Q, Wen WQ, et al. Genetic polymorphism of cytochrome P450-1B1 and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2000 Feb;9(2):147-50.

Zhu BT, Conney AH. Functional role of estrogen metabolism in target cells: review and perspectives. Carcinogenesis 1998 Jan;19(1):1-27.