University of Alberta

Genotyping the CYP*4 Allele by Allele Specific PCR and Melt Curve Analysis

by

Hashim Khan Zuhair, Mohammad

A thesis submitted to the Faculty of Graduate Studies and Research in Partial fulfillment of the requirements for the degree of

Master of Science

in

Medical Sciences – Laboratory Medicine and Pathology

Spring 2012

Edmonton, Alberta

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-87863-7

Our file Notre référence
ISBN: 978-0-494-87863-7

Abstract

Tamoxifen is one of the most effective drugs for treating breast cancer. It is metabolized by CYP2D6. CYP2D6*4, the most common allelic variant of CYP2D6 in Caucasian population is associated with the single nucleotide polymorphism rs3892097 (1846G>A), which results in a poor metabolizer phenotype for tamoxifen. Therefore, genotyping allele CYP*4 in patients will help in making better therapeutic decisions. The objective of the thesis is to develop a clinically feasible diagnostic test to detect the SNP rs3892097 (G>A), which is challenged by the presence of highly homologous pseudogenes CYP2D7 and CYP2D8P near the active gene on chromosome 22q13.1. To address these challenges, we developed a tetra-allele refractory mutation detection system which allowed us to exclude the pseudogenes and produce an allele specific PCR product that can be detected by melt curve analysis. The test is feasible on the gel post system, a point of care diagnostic tool developed in our lab.

Acknowledgements

Working as a researcher during my Masters has been a very fulfilling and never-forgotten experience. In this regard, first and foremost I extend my thanks and gratitude to my supervisor and my mentor, Linda M. Pilarski for giving me the opportunity to work in her group. Linda, I could write extensively about all I have learned from you all these years. You taught me to help others and be positive. I remain thankful to you for your endless encouragement, support and for directing my research. You have been very kind and forgiving. Thank you very much for proofreading my thesis, and assisting me at committee meetings.

My sincere thanks to my co-supervisor Dr. Jason Acker, who showed me the other side of the coin – about the business aspects of an applied research. Jason, you are truly a source of inspiration to me. You were very helpful in writing my thesis and for keeping me on track with my department requirements. I will look forward to learning more from you.

I extend my sincere thanks to the supervisory committee members Dr. Raymond Lai and Dr. John Mackey for their comments and positive feedback. Without their comments I would have not have known about the challenges in my research. I am thankful to all my committee members for spending their valuable time reading my thesis and providing me with their invaluable suggestions.

I am indebted to all my lab members, especially Dr. John Crabtree for supporting me all the time, Dr. Dammika Manage for all the discussions we had and for her healthy critiques, Dr. Alexey Atrazhev, Dr. Brian Taylor and Alex Stickel for sharing their knowledge and offering advice. I wish to acknowledge Ms. Jana Lauzon, our lab manager, for the lively discussions, help

and support. Jana, you are very kind and hard working – thank you very much. I would like to thank Dr. Stephanie Yanow for helping me with her suggestions. Many thanks to my friend, Ravi Shankar Chavali for his help in and outside the lab.

I extend my sincere thanks to Ms. Cheryl Titus and Dr. Jonathan Martin for their generous help and for keeping me informed of the department's requirements.

I would like to acknowledge the funding agency, Alberta Heritage Foundation for Medical Research (AHFMR) for supporting my research work.

Lastly and most importantly, with all respect, I extend my deepest gratitude and love to my parents encouraging me to achieve my goals, for their sacrifice, for their unconditional love and for their continuous support since I born. Today my ability to achieve is a reflection of their hard work, love and sacrifice.

Table of Contents

List of Tables

List of Figures

Abbreviations

4-OH-Tam	4-hydroxy-tamoxifen
AI	Aromatase inhibitors
ASA	Allele specific amplification
ASPCR	Allele specific PCR
CYP2D6	CYP P450 2D6 drug metabolising enzyme
CYP2D7	Pseudogene
CYP2D8PP	Pseudogene
ER+	Estrogen receptor positive
FDA	Food and drug administration
MAFFT	Multiple sequence alignment program
mPCR-RETINA	An invader assay
NCBI	National center for biotechnology information
NDM	N-desmethyl-tamoxifen
RFLP	Restriction fragment length polymorphism analysis
SNPs	Single nucleotide polymorphisms
Tetra-ARMS	Tetra allele refractory mutation system PCR

Chapter 1. Introduction

1.1 Breast Cancer and tamoxifen therapy

Breast cancer is the uncontrolled growth of breast epithelial cells. It is the most common type of cancer in women [1]. One in nine women in Canada develops breast cancer by the age of 90 years with an 87% likelihood of survival after the diagnosis [2]. Familial susceptibility contributes to 10% -25% of breast cancers. The major contributors to high risk familial breast cancer are mutations in the BRCA1 and BRCA2 genes[3][4]. BRCA1 and BRCA2 mutations contribute 65%-85% and 45%-85% respectively to the lifetime risk of developing breast cancer [5]. BRCA gene mutations are also responsible for the increased number of life years lost due to early onset of the cancer. Breast cancers which are sporadic or with BRCA2 mutations are more likely to be estrogen receptor positive (ER+) [6]. The ER+ cells have estrogen receptor protein molecules that bind to the hormone estrogen, encouraging proliferation of the cancer cells. Tamoxifen is a drug that blocks the estrogen receptors from binding to estrogen and helps in the treatment of breast cancer.

Adjuvant therapy is a treatment given to patients in addition to the primary treatment e.g. after the surgery. Tamoxifen therapy remains as the standard adjuvant therapy for premenopausal and most of the postmenopausal ER+ breast cancer patients [7, 8]. The drug has shown a positive influence on breast cancer status with increased disease free survival and overall survival [9]. When

tamoxifen is administered to patients for 5 years after surgery, it halves the annual recurrence rate and the overall mortality rate by one third in both pre and postmenopausal women [10]. Tamoxifen is the only US Food and Drug Administration (FDA) approved estrogen antagonist for the prevention of breast cancer [11], and premenopausal breast cancer [12].

Although tamoxifen is the standard therapy for pre-menopausal breast cancer, Aromatase Inhibitors (AI) is also accepted as a therapy for post-menopausal breast cancer. There are two ways of treating patients with adjuvant therapy in the postmenopausal setting 1) sequencing of hormonal therapy with tamoxifen for 2-3 years followed by AI for 2-3 years [13, 14] or 2) only AI for five years[15, 16]. Compared with the administration of tamoxifen for five years, AI for five years reduces the recurrence of cancer by 13% but does not prolong the overall survival [14, 15]. However, sequencing the hormonal therapy i.e. tamoxifen for 2-3 years followed by AI for 2-3 years, reduces recurrence by 40% [13, 17] and increases overall survival[18, 19]. Therefore, tamoxifen acts as standard therapy in both premenopausal and postmenopausal settings, and has become the standard endocrine therapy for breast cancer.

1.2 *CYP2D6* gene and tamoxifen drug metabolism

1.2.1 Polymorphisms of *CYP2D6* gene

Variations in human DNA sequences can affect an individual's response to pathogens, chemicals, drugs, vaccines and other agents [20]. The focus has been on single nucleotide variations in the genes, termed single nucleotide

polymorphisms (SNPs), which are known to affect the activity of genes [21]. The presence of SNPs in an individual might influence drug metabolism and can alter the safety and efficacy of drug therapy. Screening for these genetic variants will allow treatments to be personalized based on the efficiency of drug metabolism for a given individual. Here, I focus on optimizing a detection strategy for a SNP in the gene encoding Cytochrome p450 (*CYP2D6*), a gene that contributes to the metabolism of tamoxifen.

1.2.2 Tamoxifen metabolism

CYP2D6 encodes the CYP P450 2D6 enzyme. It is expressed predominantly in the liver and participates in metabolism of approximately 25% of clinically useful drugs including many beta-blockers, antidepressants, antipsychotics, antiarrhythmic drugs, dextromethorphan, codeine and tamoxifen [22]. Polymorphisms of *CYP2D6* cause a high degree of variability in individual responses to different drugs [22]. There are over 80 alleles identified in the *CYP2D6*gene [23]. *CYP2D6*4* is the most frequent allele in Caucasians with an allelic frequency of 25% [24], and is known to affect the activity of *CYP2D6* in tamoxifen metabolism by causing the formation of truncated proteins, and thereby altering the catalytic activity of the *CYP2D6* gene product [25].

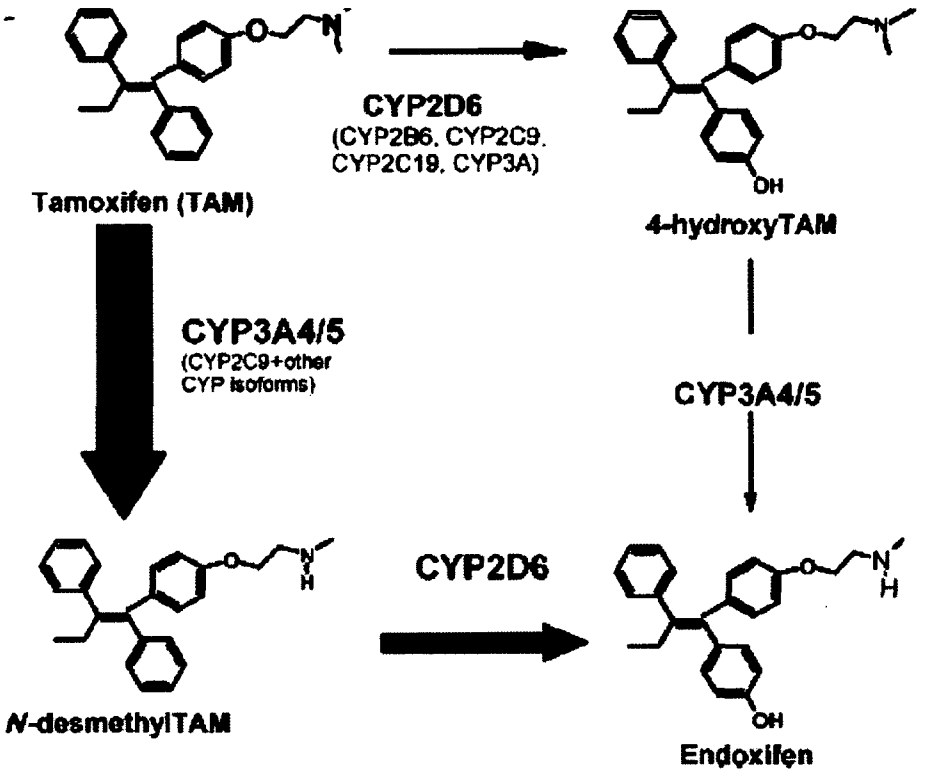
1.2.2.1 Biotransformation of tamoxifen to primary metabolites

Tamoxifen itself is not an active drug. It must be oxidised by *CYP2D6* isoforms to its primary active metabolite and secondary metabolites [26-28] (Figure 1.1). Primary metabolite include *N*-desmethyl-tamoxifen (NDM), 4-hydroxy-tamoxifen (4-OH-Tam), tamoxifen-*N*-Oxide, and α-hydroxy-tamoxifen [29-33]. NDM is the result of CYP3A4/5 mediated catalysis of tamoxifen, and it is quantitatively the major primary metabolite. It accounts for 92% of the primary metabolites [30], whereas the concentration of (4-OH-Tam) is very low. If a women receives tamoxifen therapy of 20mg/day, the steady state concentration of the metabolites in plasma are NDM =654nM, tamoxifen 362nM and 4-OH-Tam 9nM [34].

1.2.2.2 Biotransformation of tamoxifen to secondary metabolite (Endoxifen)

NDM acts as an intermediary substrate and it is bio transformed to α-hydroxy *N*-desmethyl-, *N*-didesmethyl-, and 4-hydroxy-N-desmethyl-tamoxifen (endoxifen)[30, 34, 35]. NDM biotransformation is carried out by the CYP3A subfamily. However, biotransformation to endoxifen is exclusively carried out by *CYP2D6* [30] (Figure 1.1).

Figure 1.1 Tamoxifen transformation pathways catalyzed by *CYP2D6* enzymes



The major primary metabolite is *N*-desmethyl tamoxifen which is exclusively bio-tranformed by *CYP2D6*. Hence, polymorphisms in the gene are important to determine the status of tamoxifen metabolism in breast cancer patients to reduce the toxicity of the drug and improve the therapeutic outcomes. Adapted from Boocock et al., 2002 [36].

Polymorphisms in the *CYP2D6* gene and genes acting as drug inhibitors, including antidepressants such as serotonin reuptake inhibitors, have been shown to reduce the level of endoxifen, and are associated with poor outcome in breast cancer patients treated with tamoxifen [29, 37, 38]. Polymorphisms of *CYP2D6* significantly reduce the plasma concentration of endoxifen [29, 34, 39].

The mutant form *CYP2D6*4* has altered catalytic activity; as a result endoxifen is not produced, the drug remains inactive and the patient fails to respond [25]. These variations explain the clinical variability in the plasma concentration of endoxifen, which is found to be low in women with a mutant allele at *CYP2D6*4* as compared to the women who have a wild type *CYP2D6*4* allele [25].

Individuals carrying "mutant" or minor alleles of *CYP2D6* SNPs produce reduced concentrations of endoxifen in their plasma, leading to reduced benefit from tamoxifen drug therapy [28], since the active metabolite endoxifen is 100 times more potent than tamoxifen. Depending upon the presence or absence of mutant SNP alleles, individuals differ in their ability to metabolize tamoxifen. Therefore, variants in *CYP2D6* can be used to assess a patient's ability to benefit from tamoxifen drug therapy. Patients carrying homozygous mutant alleles are poor metabolizers whereas homozygous wild type alleles have a normal phenotype. Individuals carrying different SNP alleles may not gain the same benefits from identical systemic therapies. Hence, prescribing the same tamoxifen dose for all patients may not be useful as some of them may fail to respond to therapy. Therefore, they might show reduced benefits to tamoxifen treatment. However, identification of the *CYP2D6*4* variant in a patient will allow a physician to adjust the dose of tamoxifen according to the ability of that patient to metabolize the drug, or to select alternative therapies for non-responders.

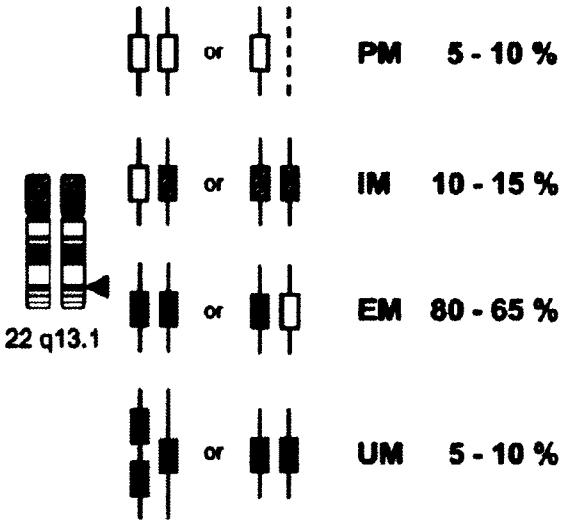
1.2.3 Importance of *CYP2D6* allelic variant *4

Cytochrome p450 (*CYP2D6*) is a very important enzyme for metabolizing tamoxifen. Multiple alleles of this gene, including insertions, single nucleotide polymorphisms, deletion of the complete gene and duplication/ multi-duplication have been identified. Mutations in *CYP2D6* allelic variants cause changes in the enzyme activity level for metabolism of various drugs. There are at least 80 allelic variants (*2 to *75) identified and designated by the Human Cytochrome P450 Allele Nomenclature Committee [40].

Variants of the gene occur at clinically significant allelic frequencies in different populations. Depending upon the type of allele or the combinations of alleles, drug metabolism capacity is mainly divided into four categories. They are ultra-rapid metabolism (UM) with increased activity, extensive metabolizer (EM) and intermediate metabolism (IM) with above normal and normal activity respectively and poor metabolism (PM) with decreased or no activity [41-44] (Figure 1.2). The alleles *3, *4, *5 and *6 are responsible for ~97% of all alleles causing a PM phenotype in Caucasians [44], hence they are considered as 'null alleles' that do not encode a functional protein. The frequency of *CYP2D6*4* allele is 22% in Swedish Caucasians and accounts for 75% of the mutant alleles [45]. In Caucasians the allele **4* occurs at the highest frequency of 17-22 %, being responsible for 70-90% of all PM's [44, 46-49]. With the presence of homozygous mutant allele **4/*4*, a patient becomes PM to tamoxifen therapy. Individuals with the PM phenotype will not derive any benefit from the normal drug dosage. At the other end of the spectrum, the UM phenotype is associated

with the adverse drug reactions due to the presence of 10- to 30-fold higher concentrations of metabolites [50]. The prevalence of the UM phenotype has been found to be 4.3% in the American Caucasian population [51]. Thus genotyping of *CYP2D6*4* could be applied to individualize drug therapy, thereby decreasing the adverse effects of tamoxifen in patients with UM and increasing dosing or altering treatment for those with PM, thereby improving the therapeutic efficacy of the drug and informing alternate therapeutic decisions. However, along with the genotyping results, other factors contributing to the pharmacokinetic variability of tamoxifen should be considered including co-medication, disease status and liver function for the proper treatment of breast cancer [52]. A prerequisite for screening the PM genotype in a clinical setting is to develop a fast, reliable and cost effective technique for the routine genotyping of patients.

Figure 1.2 Genotype to phonotype relationship



Null alleles are shown by open boxes, decreased activity alleles by hatched boxes and fully functional alleles by black boxes. The frequency of the distribution is given in the percentages for each phenotype. UM=Ultra-rapid metabolizer, EM=Extensive metabolizer, IM=Intermediate Metabolizer, PM=Poor metabolizer. *CYP2D6**4 variant is a null allele with 5-10% of the phenotype in the Caucasian population. Adapted from Zanger et al., 2004[22].

1.3 Active gene *CYP2D6* vs. pseudogenes *CYP2D7* and *CYP2D8PP*

The term pseudogene is coined in 1977 by Jacq, et al., [53]. Pseudogenes are inactivated active genes that do not encode a functional protein. They are thought to have arisen as a result of duplication events followed by acquisition of mutations in the regulatory or coding sequence, which makes the duplicated gene a non-functional gene [54], or as a result of insertion of mRNA by reverse transcriptase enzyme that lack the promoter sequence necessary for transcription.

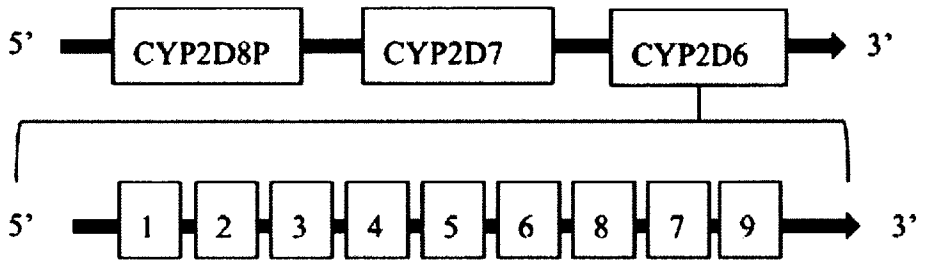
Examples of the mutations in the earlier case can be insertion or deletion in the gene sequence leading to a frame shift mutation or disarrangement of bases at splice sites or regulatory sites, which ultimately disrupt the transcription process resulting in a non-functional protein. Some pseudogenes are functional and they are transcribed to make non-coding RNAs such as the ribosomal and transfer RNAs. Very few pseudogenes are involved in regulation of gene expression [55, 56]. Therefore in general, pseudogenes are not functionally expressed genes.

The *CYP2D6* locus has pseudogenes, a single or multiple active genes or mutant gene variants. The *CYP2D6* active gene (*1 or *2 active allelic variant) is encoded downstream of two homologous genes in the order 5-*CYP2D8PP-CYP2D7-CYP2D6*–3 on chromosome 22q13 (Figure 1.3). *CYP2D8PP* and *CYP2D7* are regarded as pseudogenes by Kimura et al [57]. The pseudogenes *CYP2D7* and *CYP2D8PP* contain several gene disrupting insertions, deletions, base substitutions (transition and transversion) and termination codons within their exons that disrupt the reading frame. These pseudogenes are present up-stream of the active *CYP2D6* gene and share 97% and 96% similarity [58]. However, there are several bases where *CYP2D6* does not share homology with any pseudogene [57].

The *CYP2D6* gene consist of nine exons [57, 59] and eight intronic regions spanning 4378 bp, with a 1531 5' flanking region and a 3522bp 3' flanking region (Figure 1.3). Kimura et al found that the *CYP2D7* coding sequence contains one single inactivating mutation (an insertion T138 in the first exon causing premature translation termination), whereas *CYP2D8PP* contains

several deletions and insertions and does not have an open reading frame [57] (Table 1.1 and Table 1.2). Therefore, consideration of pseudogenes is important when designing primers to amplify regions of the *CYP2D6* active gene, so that the primers avoid pseudogene amplification.

Figure 1.3 *CYP2D6* gene cluster on chromosome 22 in 5' to 3' direction



CYP2D6 have 9 exons separated (numbered boxes) by introns (lines) [57].

Table 1.1 List of changes in *CYP2D8PP* gene as compared to *CYP2D6* active gene sequence. Adapted from Kimura et al., 1989 [57]

Exon 1	3-base deletion	Between 250 and 251
Exon 4	C→T} termination	3000
	C→A}	3002
	2-base insertion	3006–3007
Exon 5	C→T termination	3584
Exon 7	3-base deletion	Between 4177 and 4178
	C→T termination	4205
Exon 9	1-base insertion	5056

Table 1.2 Differences and similarities between *CYP2D6* active gene, *CYP2D7*, and *CYP2D8PP* pseudogenes. Adapted from Kimura et al., 1989 [57]

	Length[a] (bp)			% Similarity		
	CYP2D6	*CYP2D7*	*CYP2D8P*	*CYP 6/7*	*CYP 7/8P*	*CYP 6/8P*
Upstream	774	777		97		
		186	183		92	
	189		186			89
Exon 1	268	269	265	97	94	93
Intron 1	703	701	1,620[b]	98	90	89
Exon 2	172	172	172	95	94	91
Intron 2	550	528	546	74	78	77
Exon 3	153	553	153	98	93	92
Intron 3	88	88	88	98	91	93
Exon 4	161	161	161	98	89	91
Intron 4	433	425	449	94	85	86
Exon 5	177	177	177	99	93	92
Intron 5	190	192	186	97	84	83
Exon 6	142	142	142	94	92	96
Intron 6	207	194	204	82	87	90
Exon 7	188	188	185	98	94	95
Intron 7	454	454	449	98	91	91
Exon 8	142	142	142	99	96	96
Intron 8	98	98	96	100	97	97
Exon 9[c]	252	252		94		
		180	181		95	
	180		181			92
3' Flanking	538	528		97		

1.4 Classification of *CYP2D6* variants

Based on restriction fragment length polymorphism analysis (RFLP), the *CYP2D6* gene locus is recognized by different haplotypes (for example 42kb, 29kb, 11.5kb). The different RFLP fragment patterns arise from different mutations or duplications on the *CYP2D6* locus (Figure 1.4) [60]. Daly et al introduced a systematic nomenclature for classifying *CYP2D6* genes [61]. According to the classification, alleles that share key mutations are designated by the same allelic

number (e.g.*CYP2D6*4*) and alleles with the same key mutations but with different added mutations are recognized by letters (e.g.*CYP2D6*4*A,**4*B) [22, 61]. The *CYP2D6* allele nomenclature committee lists all the alleles and further variants of *CYP2D6* [20].

Further, the mutant gene variants are classified into alleles with increased, decreased, normal and non-functional categories [22]. For example, the null alleles of *CYP2D6* fall into the category of alleles with decreased activity giving rise to a poor metaboliser phenotype [52]. The null alleles that fail to encode a functional protein include *3,**4*,**5* and *6, and they account for ~97% of all alleles causing the PM phenotype in Caucasian populations [44]. Studies have identified **4* variant as the most common variant to be associated with PM of tamoxifen metabolism [62]. It is found that for 7%-10% of the Caucasian population, their PM phenotype is related to *CYP2D6* specific alleles [63]. The variant **4* is known to affect the activity of *CYP2D6* in tamoxifen metabolism by causing the formation of truncated proteins, and thereby affecting the catalytic activity of the gene product [64]. The non-functional enzyme is produced due to the presence of (1846G>A) at the intron3 and exon4 junction that affects splicing of mRNA (Figure 1.5) [65]. There are several SNPs in the *CYP2D6*4* allele, the most important of which is rs3892097(4180G>C) [64], since this SNP is present in 13 defined *CYP2D6*4* sub-variants (*CYP2D6*4*A,B,C,D,E,FG,H,J,K,L,M,N). Among the sub-variants, 7/13 are known to give a PM phenotype in vivo and 3/13 are known to give a PM phenotype in-vitro [40]. The SNP rs3892097 is common to all the sub-variants of *CYP2D6*4*. Therefore, detection of only the rs3892097

SNP should be sufficient to detect the presence of **4* allele. The position of this SNP on chromosome 22 is NT_011520.12:g.21915516 C>T (in minus strand).

*CYP2D6*5* is a deletion allele [66, 67] with a frequency of 2-6.9% in the Caucasian population [44, 49, 60, 63, 68] (Figure 1.6). Steen and coworkers found breakpoints in a 2.8 kb sequence flanking CYP2D6, generating deletion of the entire *CYP2D6* gene. The deletion might have resulted from the unequal crossover between homologous chromosomes or looping out of the gene on a single chromosome [69].

Figure 1.4 Example of RFLP pattern for duplication and different mutations on *CYP2D6* locus. Adapted from Ingelman-Sundberg et al., 1999 [60]

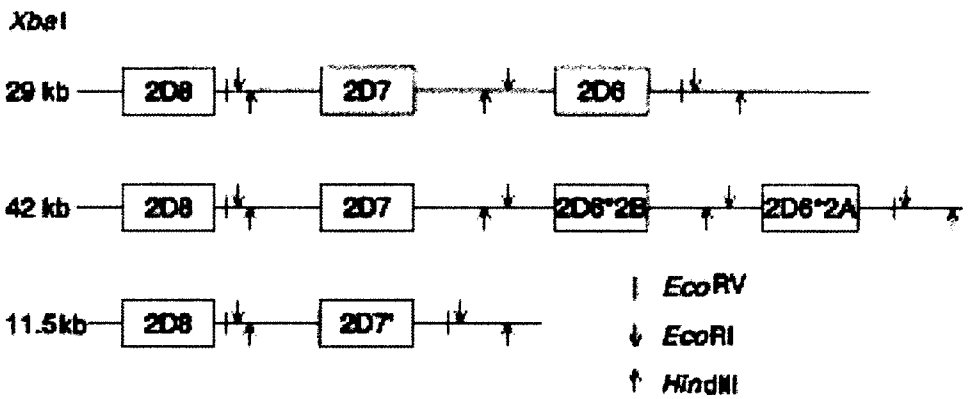
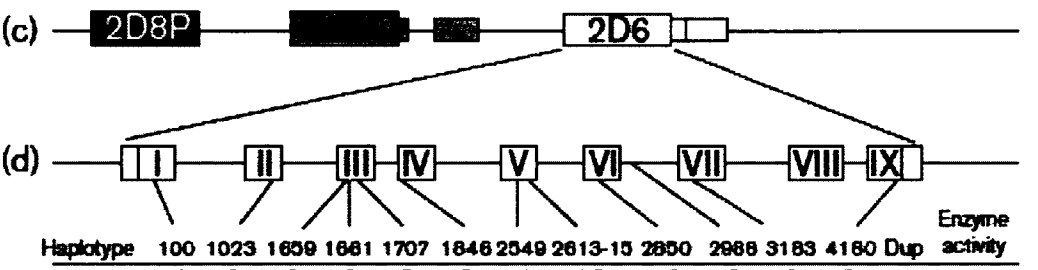
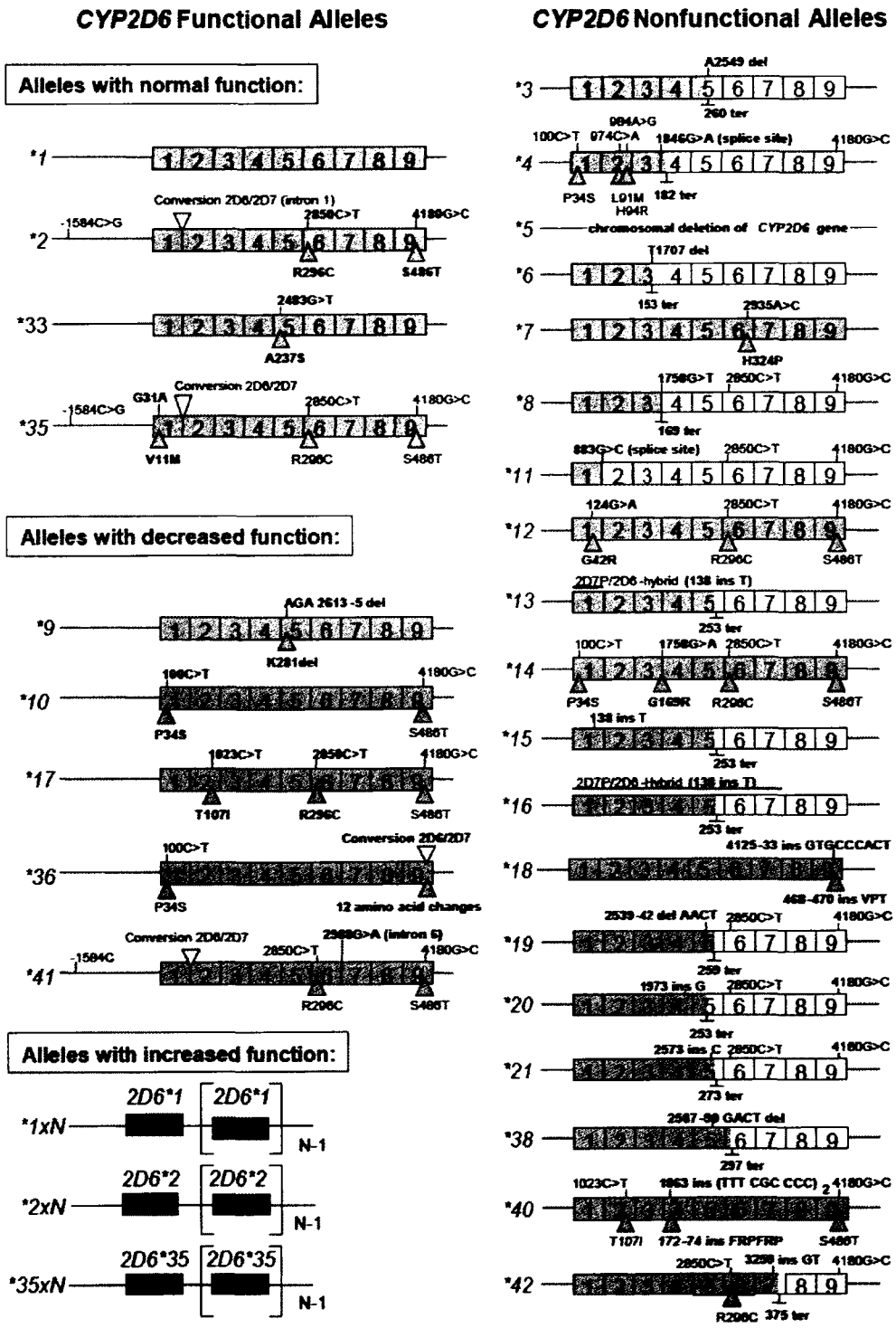


Figure 1.5 *CYP2D6*4* allele showing 1846 G to A transition in **4* and **4*B allelic variants. Adapted from Sistonen et al., 2007 [65]



Haplotype	100	1023	1659	1661	1707	1846	2549	2613-15	2850	2988	3183	4160	Dup	Enzyme activity
*1	C	C	G	G	T	G	A	AGA	C	G	G	G	–	norm
*2	C	C	G	C	T	G	A	AGA	T	G	G	C	–	norm
*3	C	C	G	G	T	G	–	AGA	C	G	G	G	–	none
*4	T	C	G	C	T	A	A	AGA	C	G	G	C	–	none
*4B	T	C	G	G	T	A	A	AGA	C	G	G	C	–	none
*5	–	–	–	–	–	–	–	–	–	–	–	–	–	none
*6	C	C	G	G	–	G	A	AGA	C	G	G	G	–	none
*9	C	C	G	G	T	G	A	–	C	G	G	G	–	decr
*10	T	C	G	C	T	G	A	AGA	C	G	G	C	–	decr
*17	C	T	G	C	T	G	A	AGA	T	G	G	C	–	decr
*29	C	C	A	C	T	G	A	AGA	T	G	A	C	–	decr
*39	C	C	G	C	T	G	A	AGA	C	G	G	C	–	norm
*41	C	C	G	C	T	G	A	AGA	T	A	G	C	–	decr
*1661	C	C	G	C	T	G	A	AGA	C	G	G	G	–	norm
*4160	C	C	G	G	T	G	A	AGA	C	G	G	C	–	norm
*1xN	C	C	G	G	T	G	A	AGA	C	G	G	G	+	incr
*2xN	C	C	G	C	T	G	A	AGA	T	G	G	C	+	incr
*4xN	T	C	G	C	T	A	A	AGA	C	G	G	C	+	none
*10xN	T	C	G	C	T	G	A	AGA	C	G	G	C	+	decr
*41xN	C	C	G	C	T	G	A	AGA	T	A	G	C	+	decr
*1661xN	C	C	G	C	T	G	A	AGA	C	G	G	G	+	incr

Figure 1.6 Structural depictions of different *CYP2D6* alleles

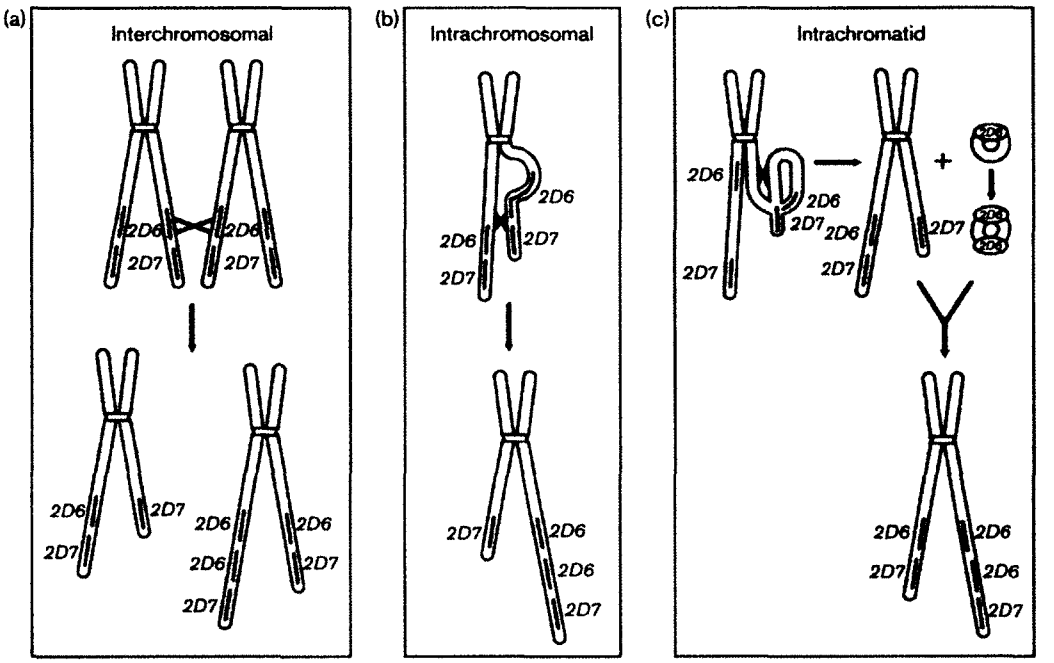


The **4* variant falls under the category of null alleles (in green box) and does not produce a functional protein, whereas *CYP2D6*5* variant is the complete deletion of the gene. Adapted from Zanger et al., 2004 [22].

1.5 Gene duplication events in *CYP2D6* locus

Gene duplication events take place at a frequency of 1.3% in Caucasians [64]. However, multi-duplication of **4* (**4*XN) resulting in a non-functional allele is a rare event in Caucasians [45]. The multi-duplication of the active gene (*CYP2D6*) or active variants (*1 or *2) will result in extremely high activity. The frequency of such duplications is more frequent in Africans (28-56%) compared to Caucasians [45]. Various models have been described for the *CYP2D6* gene duplication or multi-duplication [60, 70]. They include inter-chromosmal, intra-chromosomal or intra-chromatidic recombination mechanisms (Figure 1.7). Multiplication of the gene would result in the increased enzymatic activity, thus favouring ultra-rapid metabolism.

Figure 1.7 Recombination models describing the mechanism that might be involved in duplication or multi- duplication of *CYP2D6* gene



(a): Between two independent homologous chromosomes (b): between sister chromatids of the same chromosome (c): Intra-chromatidic, within the same chromatid arm the *CYP2D6* gene can loop out and be reinserted at the excision site or at a different locus after replication is completed. Adapted from Kramer et al., 2009 [70].

1.6 Variant *CYP2D6*4* phenotype (*1,*2=Normal)

In Caucasians, the frequencies of the haplotypes that are important for my project are in the decreasing order of *1/*1(22.8%) >*1/*2 (homozygous wild type) (14%) >*1>**4* (heterozygous) (12.1%) >*2/**4*(heterozygous) (10.5%) >**4*>**4* (homozygous mutant) (4.4%) >*2/*2 homozygous wild (3.9%) (Table 1.3) [71]. The frequency of gene duplication for the active gene is low (~1.3%) in the Caucasian population [42, 72]. However, in a rare event, subjects can have multi-duplicated copies (>2 copies) [73].

In conclusion, if the mutant allele *4 is present in a homozygous state, it can give rise to an absence of enzyme activity for tamoxifen metabolism. When it is heterozygous with the active variants (*1 or *2), the phenotype becomes one of an extensive metabolizer. If the wild type allele *1 or *2 is in the homozygous condition, it gives rise to the ultra-rapid metabolism phenotype (Table 1.4).

Table 1.3 The data shows the frequency of duplication or multi-duplication of active (*1 or *2) and/or mutant *CYP2D6* allelic variants (4*) [73]**

S.No	Haplotype	Frequency of allele in 698 subjects
1	***1/*1**	**17-19%**
2	*1/*1XN	0.4-1.6%
3	***1/*2**	**22-26%**
4	*1XN/*2	0.5-1.3%
5	*1/*2XN	0-1.6%
6	***1/*4**	**13-15%**
7	*1XN/*4	0.4.08%
8	1/*4XN	0.8-1.5%
9	***2/*2**	**9.8-12%**
10	*2/*2XN	0.8-1.6%
11	***2/*4**	**8.7-11.7%**
12	*2XN/*4	0-0.4%
13	*2/*4XN	0-0.4%
14	*4/*4XN	0.5-0.8%
15	***4/*4**	**2.1-5.4%**

The bold are the variants, the major haplotypes that will be detected in this study.

Table 1.4 Combinations of allelic variants with the mutant allele *4 that give rise to various phenotypes

Variant haplotype	Condition	Phenotype	Number of active or Mut genes
*4/*4	homozygous	PM	2
B/*4	heterozygous	IM	2
B/*4XN	heterozygous/ duplication	IM-PM	>2
A/*4	heterozygous/ duplication	EM-IM	>2
A/*4XN	heterozygous/ duplication	IM	>2
C/*4	heterozygous/ duplication	EM-IM	>2
C/*4XN	heterozygous/duplication	IM	>2
D/*4	heterozygous/duplication	EM-IM	>2
D/*4XN	heterozygous/duplication	IM	>2
A,B,C,D/A,B,C,D	homo/hetero/duplication	UM	≥2

PM = poor metabolizer, IM=intermediate metabolizer, EM = extensive metabolizer, UM =ultra-rapid metabolizer. A=*1 and *2, B = *1 or *2, C = *1XN or *2XN, D = *1XN and *2XN.

1.7 Need for a Point of Care (PoC) test for pharmacy-genetics of tamoxifen therapy

Individuals respond differently to drugs depending upon their haplotypes. Genotyping breast cancer patients for their *4 status will contribute significantly to treatment decisions informed by accurate predictions of their response to tamoxifen therapy.

Accurate genotyping is essential for this kind of personalized medicine. Currently, in a clinical setting it is difficult to routinely perform diagnostic tests that detect the presence of therapeutically important SNPs, due to increased cost of diagnosis, the need to "batch" samples resulting in extensive delays, inaccessibility of the devices and lack of skilled staff [74]. There is a need to develop an easy to operate, sensitive and low cost SNP detection method to perform diagnosis in the clinic on a "one at a time" basis for each patient. Therefore, my focus will be on developing a cost efficient and sensitive test to detect the *CYP2D6*4* SNP alleles in breast cancer patients to maximize the benefit from and efficacy of tamoxifen therapy. For genetic non responder phenotypes, a "one at a time" point of care platform could be used to inform the choice of alternate treatment at the time the patient presents. The multi-duplication of active allele and the deletion allele *5 are rare in Caucasian population. Hence, detecting the most common mutant variant *4 and the wild type alleles *1 / *2 in breast cancer patients will be sufficient to determine the status of tamoxifen metabolism, and it could serve as a prototype system to

develop multi-parameter detection strategies that also detect the other related SNPs and deletions. This pharmacogenetic test is first developed in conventional systems e.g. PCR and Lightcycler. Further, the test will be optimized for in-gel post method. The gel post system benefits from a reduced volume of the reagents (800nl master mix/ post which is one twelfth of the volume in a conventional thermocycler) which helps minimize the overall cost of diagnostics. The gel post system will also be able to carry out different target identifications in a single run (multi-parameter testing), in parallel with replicates and controls.

1.8 Conventional methods of *CYP2D6**4 detection

The *CYP2D6* gene amplification is complicated by co-amplification of pseudogenes *CYP2D7* and CYP2D8, which is resolved by pre-amplification of *CYP2D6* region followed by SNP detection by Restriction Fragment Length Polymorphism (RFLP) or Allele Specific Amplification (ASA) using the first PCR product as template. However, structural variations in the *CYP2D6* gene have led to false positive results [75, 76] and long range PCR is very time consuming [77, 78].

*Taq*Man real-time PCR [79], SNaPshot [80], AmpliChip [81], pyrosequencing [82] were developed later and give better *CYP2D6* SNP scoring. Recently PCR based invader assays (mPCR-RETINA) [83] have been shown to detect SNPs and enumerate copy numbers of the active gene. However, all these methods are still high cost, not suitable for a point of care test in the clinic and require skilled technologists for their use. They are also not suitable for

implementation on the point of care gel post platform briefly described above (and see below).

Compared to all the above methods, long range PCR of *CYP2D6* followed by Allele Specific PCR (ASPCR) appears applicable to the gel post system. It does not require any modification of oligonucleotides, thereby decreasing the overall cost of the assay. The gel post platform is able to perform thermal cycling to ensure synthesis of the *CYP2D6* specific region followed by allele specific PCR. Subsequently, the double stranded PCR product can be melted to provide melt curve verification for the ASPCR products. Therefore, I proposed use of long range PCR followed by ASPCR as a method to successfully detect the *CYP2D6**4 status in conventional systems (thermocycler for PCR and Lightcycler for melt curve analysis MCA), and still feasible for use on the gel post platform.

1.9 Gel post system

The concept of achieving PCR in a solid medium was first introduced in 1997 by Chetverin. The solid medium is a gel matrix within which DNA can be amplified. The immobilization of PCR reagents within the gel matrix prevents cross mobility and consequent cross contamination of the reagents to neighbouring gel matrices that might contain reagents for other targets. The solid medium has also been employed by other research groups [84-86]. In our lab, the gel post method was developed by Dr. Alexey Atrazhev of the AHFMR Team Microfluidics (University of Alberta) [87]. Thermal cycling is achieved on the Viriloc instrument which was developed and is being optimized by Alex Stickel

(Team Microfluidics, University of Alberta). The current prototype of the Viriloc carries out melt curve analysis with 1°C resolution (Figure 1.9).

The mold is made of two 1.1mm thick glass microscopic slides permanently bonded together, with dimensions of approximately 20x20 mm (Figure 1.8). The gel consists of acrylamide units cross-linked by bis-acrylamide, which is extensively used in electrophoresis of biomolecules, such as DNA and proteins. The gel containing all the reagents is polymerized in the form of small tablets, each measuring a volume of about 600 nanolitres. Amplification and detection strategies include seamless polymerase chain reaction (PCR) and melt curve analysis (MCA) steps within the same posts (in situ): there is no need for pumps or valves to remove the product after PCR prior to performing MCA. These gel posts on a cover slip are put in a pan containing oil and placed in the Viriloc to perform thermal cycling for PCR and MCA. A brief schematic of the Viriloc and gel posts is shown in Figure 1.9. Thermal cycling on a Peltier heating device can be employed for PCR and MCA. Simultaneously, for real time quantitative PCR, images of the gel posts can be captured with a built-in imager and stored on the computer.

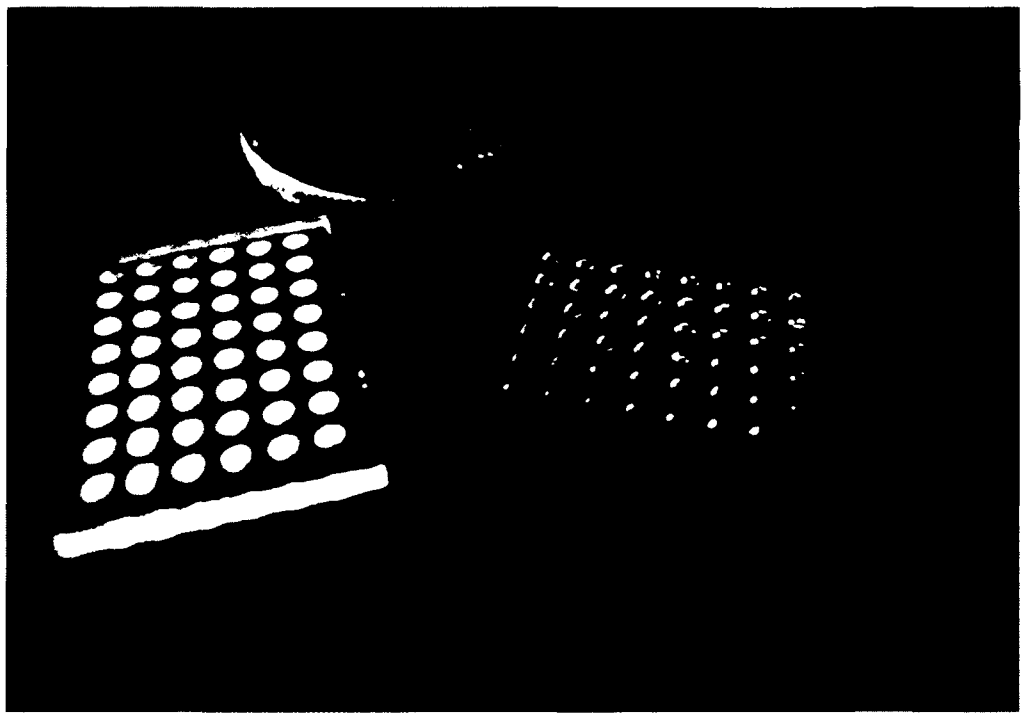
Gel post technology is ideal for performing diagnostic tests in a clinical setting on a routine basis in low-resource laboratories. The Viriloc system is a pre-programmable machine that can regulate the temperature cycling for polymerase chain reaction followed by melt curve analysis. The inexpensive components for the instrument bring down the overall cost per test. DNA amplification and analysis can be done seamlessly, removing any need for manual

intervention and reducing the risk of cross contamination. Other advantages include compactness of the system, testing many targets at a time on the same chip, and lower reagent costs.

The methodology involves photo-polymerization of a gel mix containing all the reagents, except DNA, followed by PCR and MCA, which are performed on Viriloc systems. Melt-curve analysis (MCA) can be used to verify different reaction products. MCA monitors the temperature at which double stranded DNA separates into single strands. Typically, a fluorescent intercalating dye, LC green 1, is included in the reagent mix. The intercalator binds to double stranded DNA and fluoresces when excited by a laser. The fluorescence disappears when the DNA is melted into single strands. During the process of melting, the temperature of the dsDNA is slowly increased while monitoring the fluorescence to determine the temperature where melting occurs. The fluorescence will decrease in a linear fashion as the temperature increases and will drop sharply when the DNA helix of the amplicon denatures to single strands (melts). The correct melt curve is produced only if the PCR efficiently amplified the correct product in the gel posts. If the incorrect product(s) or the primer dimers are amplified, the wrong peak or a broad melting profile is seen. If this occurs, the primer design must be refined. If the ASPCR does not amplify a product due to the mismatch of the **4* allele with the allele specific primer no melt is produced, and vice versa, with the two allele specific reactions serving as controls for each other. Therefore the gel post system could be used to genotype the **4* allele in patient samples.

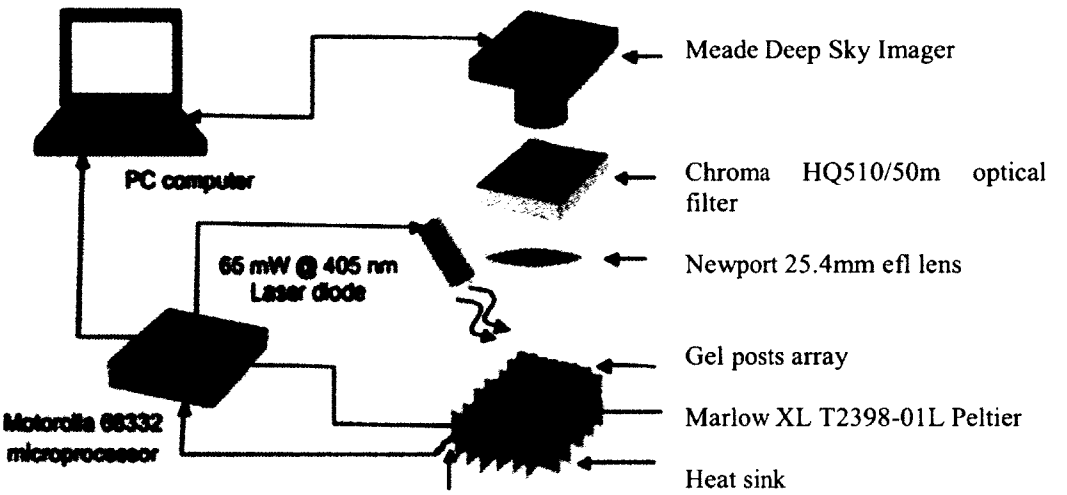
The Viriloc instrument is a prototype used to perform PCR thermocycling (Figure 1.9). During each extension phase of the PCR cycle, the laser is turned on and the camera takes a picture which is saved on the computer. During the MCA the laser is kept turned on and the images of the gel post fluorescence are taken continuously at each degree from 55 to 95 °C.

Figure 1.8 Gel post array on glass coverslip



On the right side is the cover slip carrying an array of polyacrylamide gel posts having allele specific PCR master mix within them. On the left side is the glass mold used to create the print of gel posts on the coverslip. The gel is polymerized by photo polymerisation under UV light.

Figure 1.9 Schematics of the gel post system



The printed gel posts on the coverslip are placed in the Viriloc for thermal cycling, and the raw fluorescence during the PCR and MCA can be captured by the camera and sent to the computer. Once the PCR/MCA is finished, the analysis is done using software ImageJ.xla [83].

1.10 Gel post (PoC) system vs. conventional systems

A point of care diagnostic system is used to perform test at or near the site where clinical care is delivered [88]. There are several advantages of PoC systems compared to the conventional systems such as the light cycler or pyrosequencing. PoC systems can be sent to remote areas as the systems are smaller in size and can be run easily by physicians and non-technical staff at the bed site. PoC systems do not need a permanent dedicated lab space, and the entire process of sample collection, analysis and results can be obtained near the patient care point.

Therefore, it is possible to avoid the time delays in obtaining the results making the genetic test more rapid and accessible. PoC systems reduce the risk of contamination that can occur when the samples are delivered to the testing facility as seen in conventional setting. Conventional systems need batch processing of many samples at a time to reduce the cost and labor of the test, but there is no batch processing required in the gel post setting. Each patient can be tested "one at a time" for many different targets; as a result physicians do not need to wait for longer times to obtain the result. Hence it accelerates the health care delivery and recovery times in patients. Gel post uses less reagents, space and labour, thereby reducing the overall cost of the test. Gel post testing can improve efficiency of health care by improving patients flow through busy clinics and emergency departments due to decrease waiting times.

1.11 Allele Specific Polymerase Chain Reaction (ASPCR)

The polymerase chain reaction enables amplification of target DNA by employing the polymerisation activity of the enzyme *Taq* Polymerase [89, 90]. Allele specific PCR (ASPCR) is a PCR strategy in which PCR is only possible when there is a complete match between the primer and the target DNA. The concept of primer extension for detecting single nucleotide changes was proposed by Sokolov in 1989 [91] to detect single nucleotide changes in the cystic fibrosis gene. In the same year, the concept was further extended to genotype Apo lipoprotein E [92] and diagnose genetic diseases such as haemophilia B and cystic fibrosis [93]. The match or mismatch of the primer-DNA duplex occurs during the annealing stage of PCR at the 3' end of the primer. If there is a match, preferential amplification of the matched target DNA occurs. With proper optimization, this method can discriminate a single base difference in the DNA, in our case a SNP. The amplified DNA is then detected using MCA, analyzing the temperature at which the PCR product dissociates into single strands. The mismatch primer-DNA complex is not amplified during PCR, and hence no product melt curve is generated (Figure 1.10).

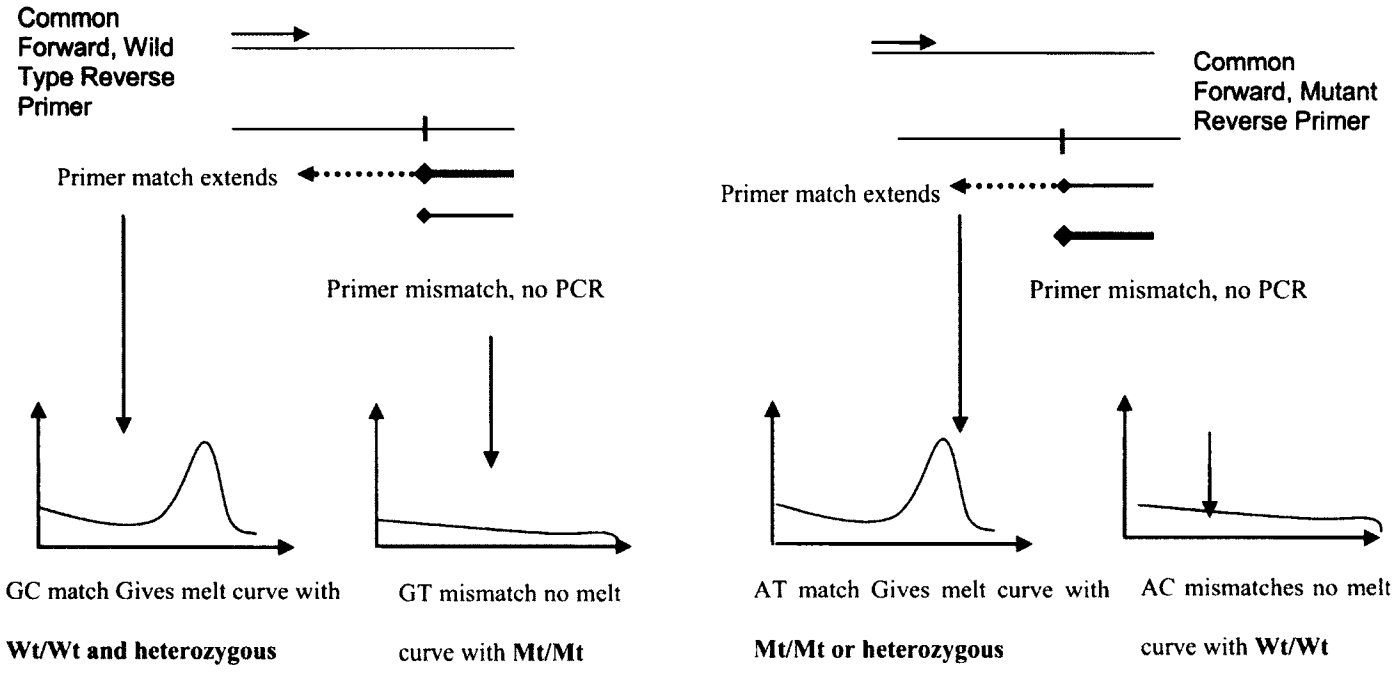
The primers can be designed manually or by using many online tools such as NCBI primer BLAST tool. Different primers can be designed and evaluated for efficiency based on the position of the primer around the SNP region, the length of the primer itself, and by the introduction of different deliberate mismatches at the penultimate (-1) or anti-penultimate (-2) position from the SNP site. The match and mismatch behaviour of the 3' binding site in the primer has revealed

some of the best matches and mismatch combinations, aided by analysis of kinetic studies of the reaction [94-99].

Based on the published literature detailing the probability of primer-DNA binding [100], DNA replication rate [101, 102] and amplification efficiency [103-107], I concluded that purine-purine or pyramidine-pyramidine mismatches in a primer-DNA duplex are most efficient for ASPCR.

After the test is developed, the results on the gel post system will be compared to the results on the conventional system. The conventional system includes ASPCR on a commercial thermocycler, PCR product MCA on the Lightcycler, and confirmation of the correct PCR product size by gel electrophoresis. The gel post system involves a seamless PCR reaction followed by MCA on the same chip without any manual intervention.

Figure 1.10 Allele specific primers are specific to wild/ mutant allele at their 3' end



If there is a match the PCR product extends and if there is a mismatch PCR does not happen. If PCR happens, the product is confirmed by MCA. Left panel shows the ASPCR with wild type primer and right with the mutant specific primer. The green primer shows the match and the red shows the mismatch at their 3' end.

1.11.1 Justification for use of allele specific PCR

Allele specific PCR primers do not need any modification to the nitrogen bases of the oligonucleotide, thus reducing the overall primer synthesis cost. The mismatch in this case will not produce the PCR product at all; as a result the presence or absence of the SNP is detected by the presence or absence of the PCR product. Conventionally, hybridisation probe based methods rely on the difference in the melting temperature of the hybridisation probe with the match or mismatch SNP site. In the hybridisation probe approach the differences between melting temperatures is low (1-3°C). Hence it is not a favourable option on prototype systems like Viriloc where achieving stable thermal calibration is difficult. However, the downside of the ASPCR approach is the need for strict PCR conditions and multiple optimization steps to make the same primers work for homozygous match, mismatch, and, in particular for the heterozygous condition in which the available match DNA concentration is depleted by half the homozygous match DNA concentration.

1.11.2 Steps in allele specific PCR

ASPCR thermal cycling comprises of three steps: (1) Denaturation: The duplex DNA is separated into single strands (2) Annealing: The ASPCR common forward binds to a location upstream to the SNP and allele specific reverse primers hybridizes at its 3' end to the SNP site (3) Extension: The reverse primer extends or fails to extend depending upon match or mismatch respectively at its 3' binding site. The reaction requires dexoynucleotide triphosphates (dNTPs) for the

synthesis of the new strand by the enzyme *Taq* polymerase. The enzyme employed in an ASPCR reaction should not have 3'-5' exonuclease activity in order to prevent it from degrading the mismatches of the allele specific primer-DNA duplex at its 3' end. Magnesium divalent ions act as a co-factor for *Taq* polymerase during the reaction. PCR buffer also helps in increasing the stability and optimum activity of polymerase enzyme.

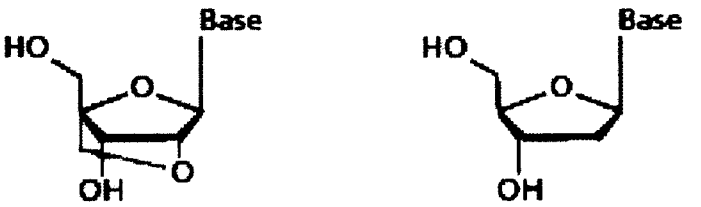
1.12 Locked Nucleic Acid (LNA) probe

When the target DNA sequence is rich in guanine and cytosine bases, the thermal stability or Tm of the product is very high. When such a sequence is amplified, a probe can bind the SNP location to discriminate between the SNP alleles. As the GC rich sequence is thermally stable, the difference between SNP allele is generally less (1-2°C) when a probe-DNA duplex is melted. My target DNA is a good example of a GC rich sequence. For such a situation, instead of an unmodified oligonucleotide probe, a "Locked Nucleic Acid" probe gives better resolution of the Tm melt curve for the SNP alleles. The Tm difference between SNPs ranges between 5-35°C [108].

LNA modified nitrogen bases were synthesised for the first time by Koshkin in 1997 [109], followed by other groups [110]. Koshkin evaluated its nucleic acid recognition potential when LNA bases are present in the DNA duplex, and they obtained an increase in thermal stability (3-8 °C) in the presence of the modified bases. LNA bases are modified ribose nucleic acid monomers (Exiqon proprietary technology) (Figure 1.11). Such monomers are added to an

oligonucleotide probe sequence to increase the thermal stability by increasing the hybridisation efficiency. The modification is the 2'-O-4'-C methylene bridge in any monomer within an oligonucleotide. The greater the number of LNA bases present, the more thermal stability is seen for the probe-DNA duplex (high Tm) [108, 111-121].

Figure 1.11 LNA monomer



LNA Monomer DNA Monomer

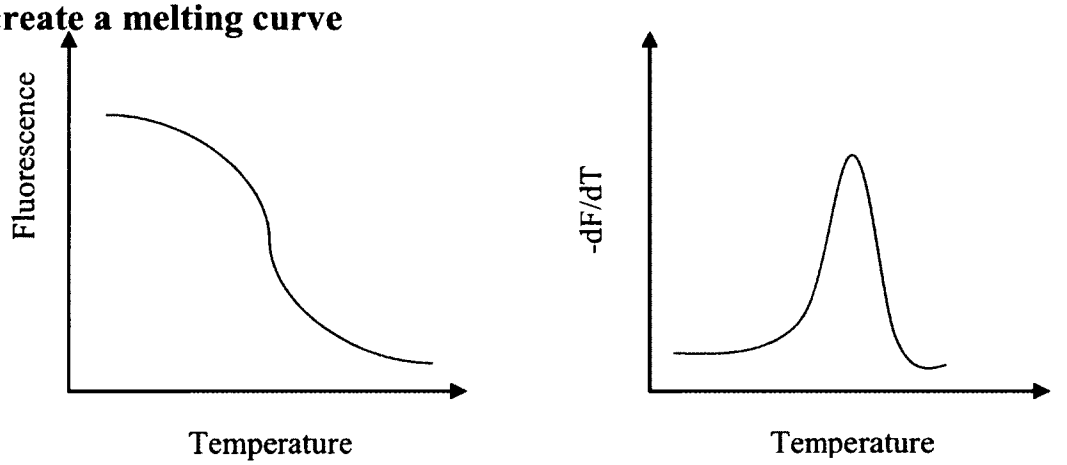
The modified RNA monomer (2'-O-4'-C methylene bridge) is an LNA base and they are incorporated in an oligonucleotide probe to increase the hybridisation interaction in probe-DNA duplex, increase the thermal stability and the discrimination efficiency between the SNP alleles during melting.

1.13 Melt Curve Analysis (MCA)

MCA assesses temperature dependent dissociation of an oligonucleotide duplex (DNA-DNA duplex or probe-DNA duplex). Every DNA duplex has specific melting temperature where more than 50% of the double stranded DNA is melted into single strands. This melting temperature depends upon the GC% of the sequence, its length and the duplex complementarity [122].

DNA duplex thermal dissociation can be measured using an intercalating fluorophore such as LC Green I. This dye is hydrophobic in nature and intercalates the minor binding grooves of the duplex DNA. The strength of the fluorescence of this dye depends upon the number of molecules of the dye bound to the duplex DNA. When the duplex is melted, the dye is released and as a result fluorescence is strongly reduced. Every duplex DNA has a certain melting temperature where more than 50% of the strands dissociate, which contributes to the large reduction in the fluorescence. The graph of the negative differential of the fluorescence with respect to temperature gives a melting peak at the specific melting temperature of the oligonucleotide duplex [123-129].

Figure 1.12 Correlation between fluorescence and temperature to create a melting curve



The left graph is the raw fluorescence values of the LC Green dye drawn at various temperatures during melting of the duplex. The right graph is the negative derivative of the raw fluorescence values again temperature. The specific melting temperature of the duplex is recognized by a sharp decrease (left graph) as a result creating a melting peak (right graph).

1.14 Hypothesis

That a "one at a time" clinically feasible diagnostic test can be developed to detect rs3892097 SNP of *CYP2D6* gene by excluding *CYP2D6* pseudogenes and amplifying only the active gene with SNP detection by allele specific PCR and MCA.

1.15 Objectives

1. Development of ASPCR in plasmid DNA on a conventional system
2. Development of ASPCR in plasmid DNA on the gel post system
3. Development of ASPCR in genomic DNA
    3.1. Population screening for
        3.1.1. Single gene, duplication and deletion
        3.1.2. SNP status
    3.2. Allele specific PCR development
        3.2.1. Excluding pseudogene strategy
        3.2.2. Addition of CYP*4 detection strategy to pseudogene exclusion within the same reaction
            3.2.2.1. Confirming the SNP detection strategy with plasmid DNA
            3.2.2.2. Performing ASPCR with pseudogene exclusion on genomic DNA
4. Batch testing known pure genomic samples for ASPCR on conventional PCR

1.16 Criteria for the assay development

Allele specific PCR demands for using the wild type and reverse primers in two separate reactions tubes (or gel posts) for the same sample in order to distinguish between homozygous wild, mutant and heterozygous types. For the homozygous wild or mutant condition, template will be amplified by PCR only in the tube containing wild type primer or mutant primer, and the template will be amplified by PCR in both the tubes for a heterozygous sample. Two negative controls (no-template controls, each carrying wild type or mutant primer but no DNA) help in identifying the false positive results. Likewise, two positive controls are needed (wild type or mutant primer with match DNA) to confirm the allele specific PCR.

CYP2D6 is highly homologous (~97%) to the pseudogenes CYP2D7 and CYP2D8P. Currently CYP2D6 is genotyped for its SNPs by amplifying a large PCR product where primers bind at the intergenic sequence specific to CYP2D6 and avoids pseudogene amplification. The result of such PCR is a long range PCR yielding a product size of 4.7kb. Such a large PCR product is not suitable for gel post application since the amplification in the gel matrix is limited by the porosity of the gel (concentration of the gel) [74]. Therefore, the criteria for assay development includes

1. The assay design should be applicable to the gel post system i.e. product size should be smaller in order to be detected, since gel matrix does not allow for the amplification of large PCR products.

2. Pseudogenes (CYP2D7, CYP2D8PP), which are highly homologous to active gene (~97%) should not contaminate the PCR reaction (primers should only target the active gene CYP2D6)
    1.1. To exclude the pseudogenes first few cycles should be performed with CYP2D6 specific primers with the least possible PCR product size, so that the amplification size is suitable for the gel post application.
    1.2. The ASPCR product has to be of the minimum size (100-500bp) in order to identify the product by MCA using LC Green dye.
    1.3. To reduce the size of the detectable PCR product, inner primer pair should be allele specific
3. Inner primer pair should reach melting temperature when outer primers are engaged in the PCR reaction so that inner primers do not interfere with the outer primer
4. PCR should happen in a nested format, so that the inner ASPCR primers can target only the CYP2D6 gene amplified during the first PCR cycles.
5. ASPCR should be performed in two different tubes for each sample. Each tube or gel post carries wild type or mutant allele specific primer pair. Both the reactions should be carried out at a time on the same gel post chip. Therefore, it is necessary that allele specific primers have similar annealing temperature, since only one annealing temperature is assigned during the PCR thermocycling.

1.17 Criteria for sample selection

28 genomic samples were genotyped to provide a known set of samples for *4 status. The assay is concerned about identifying the *4 allele as a control for test development, hence donor information is not required. The only criterion for sample size is to type a sufficient number to ensure obtaining samples of all possible genotypes (wild type homozygous, mutant homozygous and heterozygous).

1.18 Validation of the CYP2D6*4 test on gel post system

CYP2D6*4 testing is a pharmacogenetic test that seeks to promote a favorable response and prevent adverse reactions in an individual to tamoxifen based on genetic predisposition. To assess the usefulness and accuracy of CYP2D6*4 testing, four validation criteria can be considered. They are:

1. Analytical sensitivity to assess if the developed test can efficiently detect the *4 allele present in the sample
2. Analytical specificity to measure if the test is negative when the *4 allele is present in wild type form
3. Clinical validity to measure the usefulness of the analytically valid test result to diagnose patient's response to tamoxifen. It includes predictive value which can be positive or negative. Positive predictive value shows that the individual does not respond to tamoxifen treatment, whereas the negative predictive value identifies a patient's ability to metabolise tamoxifen efficiently. A CYP2D6*4 test should have a confirmed predictive value in the population to

confirm the test results. Such a predictive pharmacogenetic test is limited by incomplete penetrance (some patients who carry mutations respond to tamoxifen) and genetic heterogeneity (other mutations in CYP2D6 or other genes of CYP family are associated at similar or different levels with patient's response to tamoxifen). To provide the test with the confirmed positive predictive value, sample size should be sufficiently large to determine the frequency of the allele. It is necessary to enhance the confidence in the frequency by adding the samples as they become available in future.

4. Utility of the CYP*4 test to determine the usefulness of the test result to the patient and the clinical decisions. CYP*4 test is useful to the patients as they have other treatment options like aromatase inhibitor therapy.

References

[1] Kamangar, F., Dores, G.M., Anderson,W.F., "Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world," *Journal of Clinical Oncology,* vol. 24, pp. 2137-2150, 2006.

[2] http://www.cancer.ca

[3] Wooster, R.A , Bignell, G.a , Lancaster, J.e , Swift, S.c d , Seal, S.a , Mangion, J.a , Collins, N.a , Gregory, S.f , Gumbs, C.g , Micklem, G.f , Barfoot, R.a , Hamoudl, R.a , Patel, S.a , Rice, C.f , Biggs, P.a , Hashim, Y.a , Smith, A.c d , Connor, F.c d , Arason, A.h , Gudmundsson, J.h , Ficenec, D.h r , Kelsell, D.i , Ford, D.b , Tonin, P.j , Bishop, D.T.k , Spurr, N.K.i , Ponder, B.A.J.l , Eeles, R.a , Peto, J.b , Devilee, P.m , Cornelisse, C.m , Lynch, H.o , Narod, S.j q , Lenoir, G.n , Egilsson, V.h , Barkadottir, R.B.h , Easton, D.F.p , Bentley, D.R.f , Futreal, P.A.g , Ashworth, A.c d , Stratton,M.R.a, "Identification of the breast cancer susceptibility gene BRCA2," *Nature,* vol. 378, pp. 789-792, 1995.

[4] Miki, Y.a , Swensen, J.a , Shattuck-Eidens, D.b , Futreal, P.A.c , Harshman, K.b , Tavtigian, S.b , Liu, Q.b , Cochran, C.c , Bennett, L.M.c , Ding, W.b , Bell, R.b , Rosenthal, J.b , Hussey, C.b , Tran, T.b , McClure, M.b , Frye, C.b , Hattier, T.b , Phelps, R.b , Haugen-Strano, A.c , Katcher, H.b , Yakumo, K.a , Gholami, Z.b , Shaffer, D.b , Stone, S.b , Bayer, S.b , Wray, C.b , Bogden, R.b , Dayananth, P.b , Ward, J.d , Tonin, P.e , Narod, S.e , Bristow, P.K.f , Norris, F.H.f , Helvering, L.f , Morrison, P.f , Rosteck, P.f , Lai, M.f , Barrett, J.C.c , Lewis, C.a , Neuhausen, S.a , Cannon-Albright, L.d , Goldgar, D.a , Wiseman, R.c , Kamb, A.b , Skolnick,M.H.g, "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1," *Science,* vol. 266, pp. 66-71, 1994.

[5] Wooster, R.a , Weber,B.L.b, "Breast and ovarian cancer," *N. Engl. J. Med.,* vol. 348, pp. 2339-2347, 2003.

[6] Lakhani, S.R., Van De Vijver, M.J., Jacquemier, J., Anderson, T.J., Osin, P.P., McGuffog, L., Easton,D.F., "The pathology of familial breast cancer: Predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2," *Journal of Clinical Oncology,* vol. 20, pp. 2310-2318, 2002.

[7] Cronin-Fenton, D.P., Lash,T.L., "Clinical epidemiology and pharmacology of CYP2D6 inhibition related to breast cancer outcomes," *Expert Review of Clinical Pharmacology,* vol. 4, pp. 363-377, 2011.

[8] Galanina, N., Bossuyt, V., Harris,L.N., "Molecular predictors of response to therapy for breast cancer," *Cancer Journal,* vol. 17, pp. 96-103, 2011.

[9] Henderson, I.C., Mouridsen, H., Abe, O., Abeloff, M., Ahmann, D., Andersen, K., Baum, M., Bianco, A.R., Boccardo, F., Bonadonna, G., Buyse, M., Buzdar, A., Carbone, P., Carpenter, J., Chlebowski, R., Collins, R., Cooper, R., Crowley, J., Cuzick,J., "Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women," *N. Engl. J. Med.,* vol. 319, pp. 1681-1692, 1988.

[10] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). "Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials," *Lancet,* vol. 365, pp. 1687-1717, 2005.

[11] Fisher, B.a l , Costantino, J.P.b , Wickerham, D.L.a , Cecchini, R.S.b , Cronin, W.M.b , Robidoux, A.c , Bevers, T.B.d , Kavanah, M.T.e , Atkins, J.N.f , Margolese, R.G.g , Runowicz, C.D.h , James, J.M.i , Ford, L.G.j , Wolmark,N.a k, "Tamoxifen for the prevention of breast cancer: Current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study," *J. Natl. Cancer Inst.,* vol. 97, pp. 1652-1662, 2005.

[12] Colleoni, M., Gelber, S., Goldhirsch, A., Aebi, S., Castiglione-Gertsch, M., Price, K.N., Coates, A.S., Gelber,R.D., "Tamoxifen after adjuvant chemotherapy for premenopausal women with lymph node-positive breast cancer: International Breast Cancer Study Group trial 13-93," *Journal of Clinical Oncology,* vol. 24, pp. 1332-1341, 2006.

[13] Coombes, R.C.a v , Hall, E.b , Gibson, L.J.a , Paridaens, R.i , Jassem, J.j , Delozier, T.k , Jones, S.E.l , Alvarez, I.m , Bertelli, G.c , Ortmann, O.n , Coates, A.S.o , Bajetta, E.p , Dodwell, D.d , Coleman, R.E.e , Fallowfield, L.J.f , Mickiewicz, E.q , Andersen, J.r , LÃ¸nning, P.E.s , Cocconi, G.t , Stewart, A.g , Stuart, N.h , Snowdon, C.F.a , Carpentieri, M.u , Massimini, G.u , Bliss,J.M.b, "A Randomized Trial of Exemestane after Two to Three Years of Tamoxifen Therapy in Postmenopausal Women with Primary Breast Cancer," *N. Engl. J. Med.,* vol. 350, pp. 1081-1092, 2004.

[14] Coates, A.S., Keshaviah, A., ThÃ¼rlimann, B., Mouridsen, H., Mauriac, L., Forbes, J.F., Paridaens, R., Castiglione-Gertsch, M., Gelber, R.D., Colleoni, M., LÃ¡ng, I., Del Mastro, L., Smith, I., Chirgwin, J., Nogaret, J.-M., Pienkowski, T., Wardley, A., Jakobsen, E.H., Price, K.N., Goldhirsch,A., "Five years of letrozole compared with tamoxifen as initial adjuvant therapy for postmenopausal women with endocrine-responsive early breast cancer: Update of study BIG 1-98," *Journal of Clinical Oncology,* vol. 25, pp. 486-492, 2007.

[15] ATAC Trialists' Group, "Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer," *Lancet,* vol. 365, pp. 60-62, 2005.

[16] Thurlimann, B.a b c , Keshaviah, A.a d , Coates, A.S.a e , Mouridsen, H.a f , Mauriac, L.a g , Forbes, J.F.a h , Paridaens, R.a i , Castiglione-Gertsch, M.a j , Gelber, R.D.a d k l , Rabaglio, M.a j , Smith, I.a m n , Wardly, A.a o , Price, K.N.a d l , Goldhirsch,A.a p q, "A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer," *N. Engl. J. Med.,* vol. 353, pp. 2747-2757, 2005.

[17] Jakesz, R.a , Jonat, W.b , Gnant, M.a , Mittlboeck, M.a , Greil, R.a , Tausch, C.a , Hilfrich, J.b , Kwasny, W.a , Menzel, C.a , Samonigg, H.a , Seifert, M.a , Gademann, G.b , Kaufmann,M.b, "Switching of postmenopausal women with endocrine-responsive early breast cancer to anastrozole after 2 years' adjuvant tamoxifen: Combined results of ABCSG trial 8 and ARNO 95 trial," *Lancet,* vol. 366, pp. 455-462, 2005.

[18] Jonat, W.a , Gnant, M.b , Boccardo, F.c , Kaufmann, M.d , Rubagotti, A.c , Zuna, I.e , Greenwood, M.f , Jakesz,R.b, "Effectiveness of switching from adjuvant tamoxifen to anastrozole in postmenopausal women with hormone-sensitive early-stage breast cancer: a meta-analysis," *Lancet Oncology,* vol. 7, pp. 991-996, 2006.

[19] Coombes, R.a , Kilburn, L.b , Snowdon, C.a , Paridaens, R.c , Coleman, R.d , Jones, S.e , Jassem, J.f , Van de Velde, C.g , Delozier, T.h , Alvarez, I.i , Del Mastro, L.j , Ortmann, O.k , Diedrich, K.l , Coates, A.m , Bajetta, E.n , Holmberg, S.o , Dodwell, D.p , Mickiewicz, E.q , Andersen, J.r , LÃ¸nning, P.s , Cocconi, G.t , Forbes, J.u , Castiglione, M.m , Stuart, N.v , Stewart, A.w , Fallowfield, L.x , Bertelli, G.y , Hall, E.b , Bogle, R.z , Carpentieri, M.aa , Colajori, E.aa , Subar, M.ab , Ireland, E.b , Bliss,J.b, "Survival and safety of exemestane versus tamoxifen after 2-3 years' tamoxifen treatment (Intergroup Exemestane Study): a randomised controlled trial," *Lancet,* vol. 369, pp. 559-570, 2007.

[20] C. K. Osborne, "Tamoxifen in the treatment of breast cancer," *N. Engl. J. Med.,* vol. 339, pp. 1609-1618, 1998.

[21] F. S. Collins, L. D. Brooks and A. Chakravarti, "A DNA polymorphism discovery resource for research on human genetic variation," *Genome Res.,* vol. 8, pp. 1229, 1998.

[22] U. M. Zanger, S. Raimundo and M. Eichelbaum, "Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry," *Naunyn Schmiedebergs Arch. Pharmacol.,* vol. 369, pp. 23-37, 2004.

[23] L. D. Bradford, "CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants," *Pharmacogenomics,* vol. 3, pp. 229-243, 2002.

[24] M. Ingelman-Sundberg, "Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity," *The Pharmacogenomics Journal,* vol. 5, pp. 6-13, 2004.

[25] C. Rodriguez-Antona and M. Ingelman-Sundberg, "Cytochrome P450 pharmacogenetics and cancer," *Oncogene,* vol. 25, pp. 1679-1691, 2006.

[26] Lonning, P.E., Lien, E.A., Lundgren, S., Kvinnsland,S., "Clinical pharmacokinetics of endocrine agents used in advanced breast cancer," *Clin. Pharmacokinet.,* vol. 22, pp. 327-358, 1992.

[27] Crewe, H.K., Ellis, S.W., Lennard, M.S., Tucker,G.T., "Variable contribution of cytochromes P450 2D6, 2C9 and 3A4 to the 4-hydroxylation of tamoxifen by human liver microsomes," *Biochem. Pharmacol.,* vol. 53, pp. 171-178, 1997.

[28] V. Stearns and J. M. Rae, "Pharmacogenetics and breast cancer endocrine therapy: CYP2D6 as a predictive factor for tamoxifen metabolism and drug response?" *Expert Reviews in Molecular Medicine,* vol. 10, 2008.

[29] Stearns, V.a c , Johnson, M.D.a , Rae, J.b d , Morocho, A.b , Novielli, A.a , Bhargava, P.a b e , Hayes, D.F.a d , Desta, Z.b , Flockhart,D.A.b, "Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine," *J. Natl. Cancer Inst.,* vol. 95, pp. 1758-1764, 2003.

[30] Desta, Z.a c , Ward, B.A.a , Soukhova, N.V.b , Flockhart,D.A.a, "Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: Prominent roles for CYP3A and CYP2D6," *J. Pharmacol. Exp. Ther.,* vol. 310, pp. 1062-1075, 2004.

[31] Crewe, H.K.a b , Notley, L.M.a b , Wunsch, R.M.a b , Lennard, M.S.a b , Gillam,E.M.J.a b, "Metabolism of tamoxifen by recombinant human cytochrome P450 enzymes: Formation of the 4-hydroxy, 4-hydroxy and N-desmethyl metabolites and isomerization of trans-4-hydroxytamoxifen," *Drug Metab. Disposition,* vol. 30, pp. 869-874, 2002.

[32] Lee, K.-H., Ward, B.A., Desta, Z., Flockhart, D.A., Jones,D.R., "Quantification of tamoxifen and three metabolites in plasma by high-performance liquid chromatography with fluorescence detection: Application to a clinical trial," *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences,* vol. 791, pp. 245-253, 2003.

[33] Lien, E.A.a , Solheim, E.a , Ueland,P.M.b, "Distribution of tamoxifen and its metabolites in rat and human tissues during steady-state treatment," *Cancer Res.,* vol. 51, pp. 4837-4844, 1991.

[34] Jin, Y.a , Desta, Z.a , Stearns, V.d , Ward, B.a , Ho, H.a , Lee, K.-H.a , Skaar, T.a , Storniolo, A.M.a , Li, L.a , Araba, A.c , Blanchard, R.e , Nguyen, A.a , Ullmer, L.b , Hayden, J.b , Lemler, S.f , Weinshilboum, R.M.c , Rae, J.M.b , Hayes, D.F.b , Flockhart,D.A.a, "CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment," *J. Natl. Cancer Inst.*, vol. 97, pp. 30-39, 2005.

[35] Stearns, V.a d e , Beebe, K.L.b , Iyengar, M.c , Dube,E.c, "Paroxetine Controlled Release in the Treatment of Menopausal Hot Flashes: A Randomized Controlled Trial," *J. Am. Med. Assoc.*, vol. 289, pp. 2827-2834, 2003.

[36] D. J. Boocock, K. Brown, A. H. Gibbs, E. Sanchez, K. W. Turteltaub and I. N. H. White, "Identification of human CYP forms involved in the activation of tamoxifen and irreversible binding to DNA," *Carcinogenesis,* vol. 23, pp. 1897, 2002.

[37] Jin, Y.a , Desta, Z.a , Stearns, V.d , Ward, B.a , Ho, H.a , Lee, K.-H.a , Skaar, T.a , Storniolo, A.M.a , Li, L.a , Araba, A.c , Blanchard, R.e , Nguyen, A.a , Ullmer, L.b , Hayden, J.b , Lemler, S.f , Weinshilboum, R.M.c , Rae, J.M.b , Hayes, D.F.b , Flockhart,D.A.a, "CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment," *J. Natl. Cancer Inst.*, vol. 97, pp. 30-39, 2005.

[38] Goetz, M.P., Rae, J.M., Suman, V.J., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L., Flockhart, D.A., Desta, Z., Perez, E.A., Ingle,J.N., "Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes," *Journal of Clinical Oncology,* vol. 23, pp. 9312-9318, 2005.

[39] Borges, S., Desta, Z. , Li, L. , Skaar, T.C., Ward, B.A., Nguyen, A., Jin, Y., Storniolo, A.M., Nikoloff, D.M., Wu, L., Hillman, G., Hayes, D.F., Stearns, V., Flockhart,D.A., "Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: Implication for optimization of breast cancer treatment," *Clin. Pharmacol. Ther.*, vol. 80, pp. 61-74, 2006.

[40] www.cypalleles.ki.se/cyp2d6.htm

[41] Ingelman-Sundberg, M, "Pharmacogenetics of cytochrome P450 and its applications in drug therapy: The past, present and future," *Trends Pharmacol. Sci.*, vol. 25, pp. 193-200, 2004.

[42] M. Pirmohamed, S. James, S. Meakin, C. Green, A. K. Scott, T. J. Walley, K. Farrar, B. K. Park and A. M. Breckenridge, "Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18 820 patients," *Br. Med. J.*, vol. 329, pp. 15-19, 2004.

[43] M. Ingelman-Sundberg and C. Rodriguez-Antona, "Pharmacogenetics of drug-metabolizing enzymes: Implications for a safer and more effective drug therapy," *Philos. Trans. R. Soc. B Biol. Sci.,* vol. 360, pp. 1563-1570, 2005.

[44] C. Sachse, J. Brockmöller, S. Bauer and I. Roots, "Cytochrome P450 2D6 variants in a Caucasian population: Allele frequencies and phenotypic consequences," *Am. J. Hum. Genet.,* vol. 60, pp. 284-295, 1997.

[45] A. Gaedigk, L. Ndjountche, K. Divakaran, L. D. A. Bradford, I. Zineh, T. Oberlander, D. Brousseau, D. McCarver, J. Johnson and S. Alander, "Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events," *Clinical Pharmacology & Therapeutics,* vol. 81, pp. 242-251, 2007.

[46] M. Kagimoto, M. Heim, K. Kagimoto, T. Zeugin and U. Meyer, "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.,* vol. 265, pp. 17209, 1990.

[47] D. Marez, M. Legrand, N. Sabbagh, J. Lo Guidice, C. Spire, J. Lafitte, U. A. Meyer and F. Broly, "Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: Characterization of 48 mutations and 53 alleles, their frequencies and evolution," *Pharmacogenetics,* vol. 7, pp. 193-202, 1997.

[48] E. Griese, U. M. Zanger, U. Brudermanns, A. Gaedigk, G. Mikus, K. Mörike, T. Stüven and M. Eichelbaum, "Assessment of the predictive power of genotypes for the in-vivo catalytic function of CYP2D6 in a German population," *Pharmacogenetics,* vol. 8, pp. 15-26, 1998.

[49] S. Raimundo, C. Toscano, K. Klein, J. Fischer, E. Griese, M. Eichelbaum, M. Schwab and U. M. Zanger, "A novel intronic mutation, 2988G>A, with high predictivity for impaired function of cytochrome P450 2D6 in white subjects," *Clin. Pharmacol. Ther.,* vol. 76, pp. 128-138, 2004.

[50] P. Dalén, C. Frengell, M. Dahl and F. Sjöqvist, "Quick onset of severe abdominal pain after codeine in an ultrarapid metabolizer of debrisoquine," *Ther. Drug Monit.,* vol. 19, pp. 543-544, 1997.

[51] S. J. London, A. K. Daly, J. B. S. Leathart, W. C. Navidi, C. C. Carpenter and J. R. Idle, "Genetic polymorphism of CYP2D6 and lung cancer risk in African-Americans and Caucasians in Los Angeles County," *Carcinogenesis,* vol. 18, pp. 1203-1214, 1997.

[52] S. F. Zhou, "Polymorphism of human cytochrome p450 2D6 and its clinical significance: Part i," *Clin. Pharmacokinet.,* vol. 48, pp. 689-723, 2009.

[53] C. Jacq, J. R. Miller and G. G. Brownlee, "A pseudogene structure in 5S DNA of Xenopus laevis," *Cell,* vol. 12, pp. 109-120, 1977.

[54] E. F. Vanin, "Processed pseudogenes: characteristics and evolution." *Annu. Rev. Genet.,* vol. 19, pp. 253-272, 1985.

[55] E. S. Balakirev and F. J. Ayala, "Pseudogenes: Are They "Junk" or Functional DNA?" *Annu. Rev. Genet.,* vol. 37, pp. 123-151, 2003.

[56] R. Baertsch, M. Diekhans, W. J. James, D. Haussler and J. Brosius, "Retrocopy contributions to the evolution of the human genome," *BMC Genomics,* vol. 9, 2008.

[57] S. Kimura, M. Umeno, R. C. Skoda, U. A. Meyer and F. J. Gonzalez, "The human debrisoquine 4-hydroxylase (CYP2D) locus: Sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene," *Am. J. Hum. Genet.,* vol. 45, pp. 889-904, 1989.

[58] M. H. Heim and U. A. Meyer, "Evolution of a highly polymorphic human cytochrome P450 gene cluster: CYP2D6," *Genomics,* vol. 14, pp. 49-58, 1992.

[59] D. W. Nebert, D. R. Nelson, M. Adesnik, M. J. Coon, R. W. Estabrook, F. J. Gonzalez, F. P. Guengerich, I. C. Gunsalus, E. F. Johnson And B. Kemper, "The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci," *DNA,* vol. 8, pp. 1-13, 1989.

[60] E. Lundqvist, I. Johansson and M. Ingelman-Sundberg, "Genetic mechanisms for duplication and multiduplication of the human CYP2D6 gene and methods for detection of duplicated CYP2D6 genes," *Gene,* vol. 226, pp. 327-338, 1999.

[61] M. Ingelman-Sundberg, A. K. Daly, M. Oscarson and D. W. Nebert, "Human cytochrome P450 (CYP) genes: Recommendations for the nomenclature of alleles," *Pharmacogenetics,* vol. 10, pp. 91-93, 2000.

[62] W. Schroth, L. Antoniadou, P. Fritz, M. Schwab, T. Muerdter, U. M. Zanger, W. Simon, M. Eichelbaum and H. Brauch, "Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes," *J. Clin. Oncol.,* vol. 25, pp. 5187-5193, 2007.

[63] A. R. Hartman and P. Helft, "The ethics of CYP2D6 testing for patients considering tamoxifen." *Breast Cancer Res.,* vol. 9, pp. 103, 2007.

[64] A. Gaedigk, L. D. A. Bradford, S. W. Alander and J. S. Leeder, "CYP2D6* 36 gene arrangements within the CYP2D6 locus: association of CYP2D6* 36 with poor metabolizer status," *Drug Metab. Disposition,* vol. 34, pp. 563, 2006.

[65] J. Sistonen, A. Sajantila, O. Lao, J. Corander, G. Barbujani and S. Fuselli, "CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure," *Pharmacogenet. Genomics,* vol. 17, pp. 93-101, 2007.

[66] V. M. Steen, O. A. Andreassen, A. K. Daly, T. Tefre, A. Brresen, J. R. Idle and A. Gulbrandsen, "Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology," *Pharmacogenetics,* vol. 5, pp. 215-223, 1995.

[67] A. Gaedigk, M. Blum, R. Gaedigk, M. Eichelbaum and U. A. Meyer, "Deletion of the entire cytochrome P450 CYP2D6 gene as a cause of impaired drug metabolism in poor metabolizers of the debrisoquine/sparteine polymorphism," *Am. J. Hum. Genet.,* vol. 48, pp. 943-950, 1991.

[68] D. R. Nelson, D. C. Zeldin, S. M. G. Hoffman, L. J. Maltais, H. M. Wain and D. W. Nebert, "Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants," *Pharmacogenetics,* vol. 14, pp. 1-18, 2004.

[69] V. M. Steen, A. Molven, N. K. Aarskog and A. Gulbrandsen, "Homologous unequal cross-over involving a 2.8 kb direct repeat as a mechanism for the generation of allelic variants of the human cytochrome P450 CYP2D6 gene," *Hum. Mol. Genet.,* vol. 4, pp. 2251-2257, 1995.

[70] W. E. Kramer, D. L. Walker, D. J. O'Kane, D. A. Mrazek, P. K. Fisher, B. A. Dukek, J. K. Bruflat and J. L. Black, "CYP2D6: Novel genomic structures and alleles," *Pharmacogenet. Genomics,* vol. 19, pp. 813-822, 2009.

[71] P. Neafsey, G. Ginsberg, D. Hattis and B. Sonawane, "Genetic polymorphism in cytochrome P450 2D6 (CYP2D6): Population distribution of CYP2D6 activity," *J. Toxicol. Environ. Health Part B Crit. Rev.,* vol. 12, pp. 334-361, 2009.

[72] A. L. Zackrisson, B. Lindblom and J. Ahlner, "High frequency of occurrence of CYP2D6 gene duplication/multiduplication indicating ultrarapid metabolism among suicide Cases," *Clin. Pharmacol. Ther.,* vol. 88, pp. 354-359, 2010.

[73] E. Aklillu, I. Persson, L. Bertilsson, I. Johansson, F. Rodrigues and M. Ingelman-Sundberg, "Frequent distribution of ultrarapid metabolizers of debrisoquine in an ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles," *J. Pharmacol. Exp. Ther.,* vol. 278, pp. 441-446, 1996.

[74] A. Atrazhev, D. P. Manage, A. J. Stickel, H. J. Crabtree, L. M. Pilarski and J. P. Acker, "In-Gel technology for PCR genotyping and pathogen detection," *Anal. Chem.*, vol. 82, pp. 8079-8087, 2010.

[75] P. Suwannasri, W. Thongnoppakhun, P. Pramyothin, A. Assawamakin and C. Limwongse, "Combination of multiplex PCR and DHPLC-based strategy for CYP2D6 genotyping scheme in Thais," *Clin. Biochem.*, vol. 44, pp. 1144-1152, 2011.

[76] T. Fukuda, H. Maune, Y. Ikenaga, M. Naohara, K. Fukuda and J. Azuma, "Novel structure of the CYP2D6 gene that confuses genotyping for the CYP2D6*5 allele." *Drug Metabolism and Pharmacokinetics.*, vol. 20, pp. 345-350, 2005.

[77] S. H. Gan, R. Ismail, W. A. Wan Adnan and Z. Wan, "Standard CYP2D6 genotyping procedures fail for the CYP2D6*5 and duplication alleles when hair roots are used as a source of DNA," *Clinica Chimica Acta,* vol. 329, pp. 61-68, 2003.

[78] M. Nevilie, R. Selzer, B. Aizenstein, M. Maguire, K. Hogan, R. Walton, K. Welsh, B. Neri and M. de Arruda, "Characterization of cytochrome P450 2D6 alleles using the Invader system," *BioTechniques,* vol. 32, pp. S34-S43, 2002.

[79] E. Schaeffeler, M. Schwab, M. Eichelbaum and U. M. Zanger, "CYP2D6 Genotyping Strategy Based on Gene Copy Number Determination by *Taq*Man Real-Time PCR," *Hum. Mutat.*, vol. 22, pp. 476-485, 2003.

[80] S. Fuselli, I. Dupanloup, E. Frigato, F. Cruciani, R. Scozzari, P. Moral, J. Sistonen and G. Barbujani, "Molecular diversity at the CYP2D6 locus in the Mediterranean region," *European Journal of Human Genetics,* vol. 12, pp. 916-924, 2004.

[81] T. Heller, J. Kirchheiner, V. W. Armstrong, H. Luthe, M. Tzvetkov, J. Brockmöller and M. Oellerich, "AmpliChip CYP450 GeneChip®: A new gene chip that allows rapid and accurate CYP2D6 genotyping," *Ther. Drug Monit.*, vol. 28, pp. 673-677, 2006.

[82] E. Söderbäck, A. Zackrisson, B. Lindblom and A. Alderborn, "Determination of CYP2D6 gene copy number by pyrosequencing," *Clin. Chem.*, vol. 51, pp. 522-531, 2005.

[83] N. Hosono, M. Kato, K. Kiyotani, T. Mushiroda, S. Takata, H. Sato, H. Amitani, Y. Tsuchiya, K. Yamazaki, T. Tsunoda, H. Zembutsu, Y. Nakamura and M. Kubo, "CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection," *Clin. Chem.*, vol. 55, pp. 1546-1554, 2009.

[84] A. B. Chetverin and H. V. Chetverina, "Accurate diagnostics with the use of molecular colonies," *Mol. Biol.,* vol. 36, pp. 244-250, 2002.

[85] J. Montgomery, C. T. Wittwer, R. Palais and L. Zhou, "Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis," *Nat. Protoc.,* vol. 2, pp. 59-66, 2007.

[86] R. D. Mitra and G. M. Church, "In situ localized amplification and contact replication of many individual DNA molecules." *Nucleic Acids Res.,* vol. 27, 1999.

[87] A. Atrazhev, D. P. Manage, A. J. Stickel, H. J. Crabtree, L. M. Pilarski and J. P. Acker, "In-Gel technology for PCR genotyping and pathogen detection," *Anal. Chem.,* vol. 82, pp. 8079-8087, 2010.

[88] J. H. Nichols, "Point of Care Testing," *Clin. Lab. Med.,* vol. 27, pp. 893-908, 2007.

[89] K. Mullis, F. Faloona and S. Scharf, "Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction," *Cold Spring Harb. Symp. Quant. Biol.,* vol. 51, pp. 263-273, 1986.

[90] R. K. Saiki, S. Scharf and F. Faloona, "Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia," *Science,* vol. 230, pp. 1350-1354, 1985.

[91] B. P. Sokolov, "Primer extension technique for the detection of single nucleotide in genomic DNA," *Nucleic Acids Res.,* vol. 18, pp. 3671, 1990.

[92] A. Syvanen, K. Aalto-Setala, L. Harju, K. Kontula and H. Soderlund, "A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E," *Genomics,* vol. 8, pp. 684-692, 1990.

[93] M. N. Kuppuswamy, J. W. Hoffmann, C. K. Kasper, S. G. Spitzer, S. L. Groce and S. P. Bajaj, "Single nucleotide primer extension to detect genetic diseases: Experimental application to hemophilia B (factor IX) and cystic fibrosis genes," *Proc. Natl. Acad. Sci. U. S. A.,* vol. 88, pp. 1143-1147, 1991.

[94] M. Huang, N. Arnheim and M. F. Goodman, "Extension of base mispairs by *Taq* DNA polymerase: Implications for single nucleotide discrimination in PCR," *Nucleic Acids Res.,* vol. 20, pp. 4567-4573, 1992.

[95] S. Kwok, D. E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson and J. J. Sninsky, "Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies," *Nucleic Acids Res.,* vol. 18, pp. 999-1005, 1990.

[96] J. P. Day, D. Bergstrom, R. P. Hammer and F. Barany, "Nucleotide analogs facilitate base conversion with 3' mismatch primers," *Nucleic Acids Res.*, vol. 27, pp. 1810-1818, 1999.

[97] C. Christopherson, J. Sninsky and S. Kwok, "The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies," *Nucleic Acids Res.*, vol. 25, pp. 654-658, 1997.

[98] L. V. Mendelman, J. Petruska and M. F. Goodman, "Base mispair extension kinetics. Comparison of DNA polymerase α and reverse transcriptase," *J. Biol. Chem.*, vol. 265, pp. 2338-2346, 1990.

[99] L. B. Dyachenko, A. A. Chenchik, G. L. Khaspekov, A. O. Tatarenko and R. S. Bibilashvili, "Efficiency of DNA synthesis initiated by complementary and mismatched primers," *Mol. Biol.*, vol. 28, pp. 654-660, 1994.

[100] M. V. Velikanov and R. Kapral, "Polymerase chain reaction: A Markov process approach," *J. Theor. Biol.*, vol. 201, pp. 239-249, 1999.

[101] G. Stolovitzky and G. Cecchi, "Efficiency of DNA replication in the polymerase chain reaction," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, pp. 12947-12952, 1996.

[102] J. Peccoud and C. Jacob, "Statistical estimations of PCR amplification rates," *Gene Quantification,* pp. 111-128, 1998.

[103] L. Raeymaekers, "Quantitative PCR: Theoretical considerations with practical implications," *Anal. Biochem.*, vol. 214, pp. 582-585, 1993.

[104] S. Schnell and C. Mendoza, "Theoretical description of the polymerase chain reaction," *J. Theor. Biol.*, vol. 188, pp. 313-318, 1997.

[105] S. Schnell and C. Mendoza, "Closed form solution for time-dependent enzyme kinetics," *J. Theor. Biol.*, vol. 187, pp. 207-212, 1997.

[106] G. Weiss and A. Von Haeseler, "A coalescent approach to the polymerase chain reaction," *Nucleic Acids Res.*, vol. 25, pp. 3082-3087, 1997.

[107] C. M. Waterfall, R. Eisenthal and B. D. Cobb, "Kinetic characterisation of primer mismatches in allele-specific PCR: A quantitative assessment," *Biochem. Biophys. Res. Commun.*, vol. 299, pp. 715-722, 2002.

[108] P. Mouritzen, A. T. Nielsen, H. M. Pfundheller, Y. Choleva, L. Kongsbak and S. Møller, "Single nucleotide polymorphism genotyping using locked nucleic acid (LNA™)," *Expert Review of Molecular Diagnostics,* vol. 3, pp. 27-38, 2003.

[109] A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition," *Tetrahedron,* vol. 54, pp. 3607-3630, 1998.

[110] S. Obika, D. Nanbu, Y. Hari, K. Morio, Y. In, T. Ishida and T. Imanishi, "Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed c3-endo sugar puckering," *Tetrahedron Lett.,* vol. 38, pp. 8735-8738, 1997.

[111] S. Obika, D. Nanbu, Y. Hari, J. Andoh, K. Morio, T. Doi and T. Imanishi, "Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2'-O,4'-C- methyleneribonucleosides," *Tetrahedron Lett.,* vol. 39, pp. 5401-5404, 1998.

[112] A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition," *Tetrahedron,* vol. 54, pp. 3607-3630, 1998.

[113] A. A. Koshkin, P. Nielsen, M. Meldgaard, V. K. Rajwanshi, S. K. Singh and J. Wengel, "LNA (locked nucleic acid): An RNA mimic forming exceedingly stable LNA:LNA duplexes [3]," *J. Am. Chem. Soc.,* vol. 120, pp. 13252-13253, 1998.

[114] M. Petersen, C. B. Nielsen, K. E. Nielsen, G. A. Jensen, K. Bondensgaard, S. K. Singh, V. K. Rajwanshi, A. A. Koshkin, B. M. Dahl, J. Wengel and J. P. Jacobsen, "The conformations of locked nucleic acids (LNA)," *Journal of Molecular Recognition,* vol. 13, pp. 44-53, 2000.

[115] D. A. Braasch and D. R. Corey, "Locked nucleic acid (LNA): Fine-tuning the recognition of DNA and RNA," *Chemistry and Biology,* vol. 8, pp. 1-7, 2001.

[116] R. Crinelli, M. Bianchi, L. Gentilini and M. Magnani, "Design and characterization of decoy oligonucleotides containing locked nucleic acids," *Nucleic Acids Res.,* vol. 30, pp. 2435-2443, 2002.

[117] M. D. SØrensen, L. KværnØ, T. Bryld, A. E. Håkansson, B. Verbeure, G. Gaubert, P. Herdewijn and J. Wengel, "α-L-ribo-configured locked nucleic acid (α-L-LNA): Synthesis and properties," *J. Am. Chem. Soc.,* vol. 124, pp. 2164-2176, 2002.

[118] M. Petersen, K. Bondensgaard, J. Wengel and J. Peter Jacobsen, "Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids," *J. Am. Chem. Soc.,* vol. 124, pp. 5974-5982, 2002.

[119] A. N. Elayadi, D. A. Braasch and D. R. Corey, "Implications of high-affinity hybridization by locked nucleic acid oligomers for inhibition of human telomerase," *Biochemistry (N. Y.),* vol. 41, pp. 9973-9981, 2002.

[120] N. Kumar, K. E. Nielsen, S. Maiti and M. Petersen, "Triplex formation with α-L-LNA (α-L-ribo-configured locked nucleic acid)," *J. Am. Chem. Soc.,* vol. 128, pp. 14-15, 2006.

[121] S. Kauppinen, B. Vester and J. Wengel, "Locked nucleic acid: high-affinity targeting of complementary RNA for RNomics." *Handbook of Experimental Pharmacology.,* pp. 405-422, 2006.

[122] A. T. Ansevin, D. L. Vizard, B. W. Brown and J. McConathy, "High resolution thermal denaturation of DNA. I. Theoretical and practical considerations for the resolution of thermal subtransitions," *Biopolymers,* vol. 15, pp. 153-174, 1976.

[123] K. M. Ririe, R. P. Rasmussen and C. T. Wittwer, "Product differentiation by analysis of DNA melting curves during the polymerase chain reaction," *Anal. Biochem.,* vol. 245, pp. 154-160, 1997.

[124] M. J. Lay and C. T. Wittwer, "Real-time fluorescence genotyping of factor V leiden during rapid-cycle PCR," *Clin. Chem.,* vol. 43, pp. 2262-2267, 1997.

[125] C. T. Wittwer, "Making DNA melting useful," *Clin. Chem.,* vol. 56, pp. 1500-1501, 2010.

[126] E. Lyon and C. T. Wittwer, "LightCycler technology in molecular diagnostics," *Journal of Molecular Diagnostics,* vol. 11, pp. 93-101, 2009.

[127] C. N. Gundry, J. G. Vandersteen, G. H. Reed, R. J. Pryor, J. Chen and C. T. Wittwer, "Amplicon melting analysis with labeled primers: A closed-tube method for differentiating homozygotes and heterozygotes," *Clin. Chem.,* vol. 49, pp. 396-406, 2003.

[128] C. T. Wittwer, G. H. Reed, C. N. Gundry, J. G. Vandersteen and R. J. Pryor, "High-resolution genotyping by amplicon melting analysis using LCGreen," *Clin. Chem.,* vol. 49, pp. 853-860, 2003.

[129] R. H. A. M. Vossen, E. Aten, A. Roos and J. T. Den Dunnen, "High-resolution melting analysis (HRMA) - More than just sequence variant screening," *Hum. Mutat.,* vol. 30, pp. 860-866, 2009.

Chapter 2. Test development in plasmid DNA

Patients with the mutant allele CYP2D6*4 are non-responders to the drug tamoxifen. The major contributor to the mutant allele *4 phenotype is a SNP, rs3892097, which is a G to A transition at 1846 base position in the forward (5'-3') strand as published by Kimura et al. However, among the SNP detection strategies, Allele Specific PCR (ASPCR) best suits the purpose of developing the test both conventionally and later on the in gel post system. The ASPCR strategy uses simple oligonucleotides and does not require additional chemistry like dyes or fluorophore tags at the 5' or 3' ends, hence reducing the overall cost of the test.

There are two main steps for ASPCR 1) amplification of the target allele (wild/ mutant) by ASPCR 2) the detection of the PCR product by melt curve analysis. Firstly, the test was developed in a conventional system. ASPCR was carried out in a thermocycler and the product size was confirmed with gel electrophoresis. Then the ASPCR product was analysed for its melting temperature in the Roche Lightcycler.

Initially the test was developed using plasmid DNA as the target. The cloned plasmid DNA includes the SNP and its flanking region from genomic DNA. After successfully implementing the strategy with plasmid DNA, the PCR protocol with the existing primers was tested using genomic DNA. However, the primers were abandoned as they can contaminate the PCR product by annealing to the pseudogenes (CYP2D8P, CYP2D7) and amplifying them along with the

active gene. All the abandoned primers and approaches for detecting CYP2D6 are reported in the Appendix.

Here, I present the ASPCR method to detect *4 alleles. This strategy was successfully implemented using plasmid DNA both conventionally and in the gel post system. I next describe the challenges posed by pseudogenes when using genomic DNA, and the evolution of the Tetra-ARMS PCR (variant of allele specific PCR) strategy to achieve the SNP detection in conventional system.

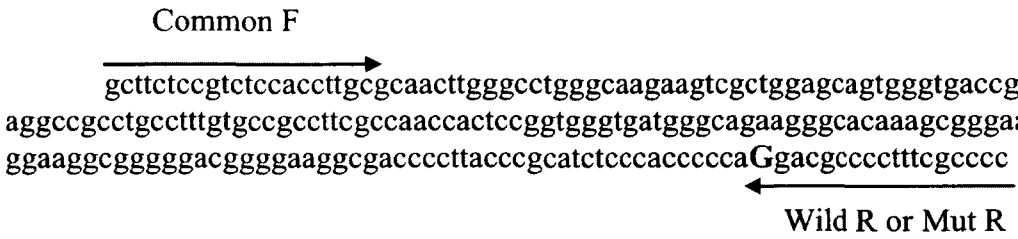
2.1 ASPCR strategy

ASPCR employed a common forward primer that was paired with either a wild type reverse or mutant reverse primer. Each primer set (wild or mutant) reaction was carried out independently in two different tubes in a thermocycler or in two different gel posts. The product size was 213 base pairs, and produced a sharp melting peak upon MCA (Figure 2.1). There was a deliberate mismatch at the penultimate base (one base behind the SNP) to increase the stringency of the PCR reaction. Hence, in the case of a mismatch at the SNP base there were in total two mismatches that completely stopped the extension of the primer during PCR. For a match primer there was still one mismatch that did not stop the amplification of match pDNA.

The ASPCR reaction was optimised for $MgCl_2$ concentration, annealing temperatures, thermal cycling parameters and reagent concentrations in both

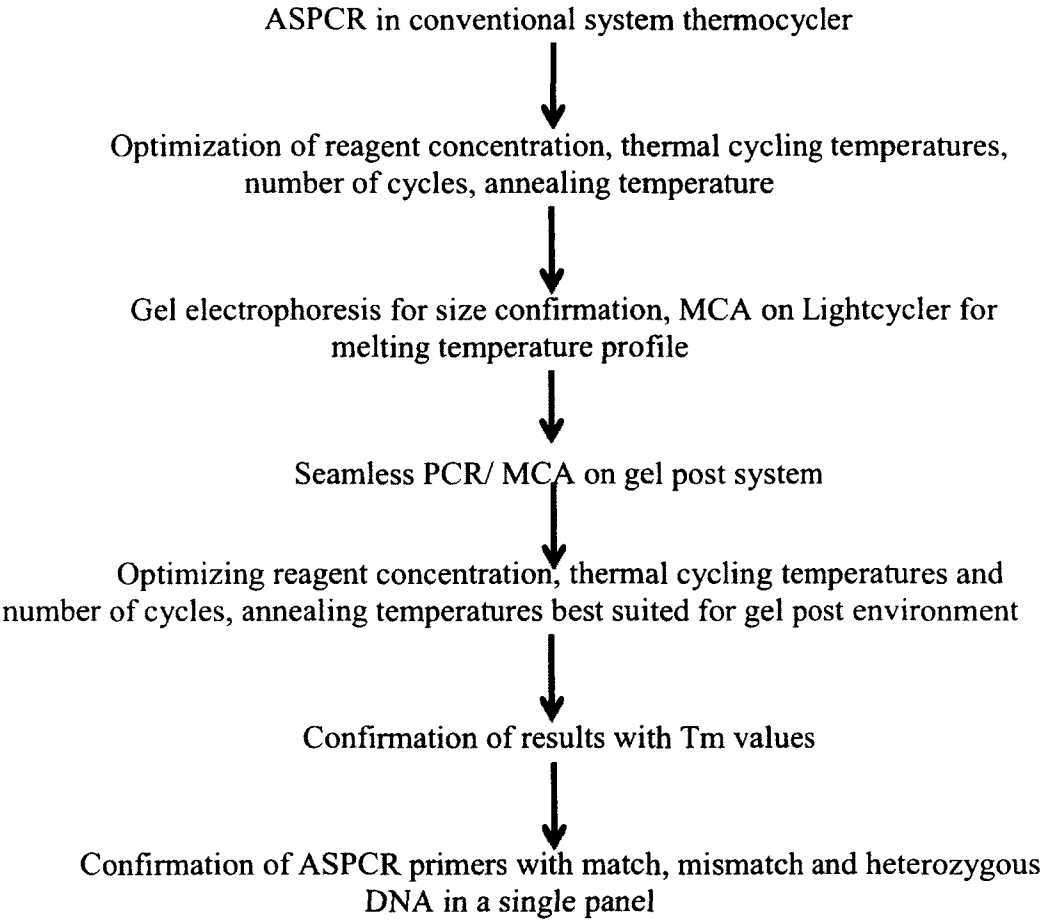
conventional and gel post systems. It is important to note that the annealing temperature of the PCR in both the reactions (wild and mutant) were optimized in the same range to allow the simultaneous runs of both reactions on the same gel post array. ASPCR using the plasmid templates was successfully optimized in both conventional and gel post systems. The experimental strategy to detect SNP in the plasmids is shown in Figure 2.2.

Figure 2.1 The ASPCR reaction in plasmid DNA



The reaction was carried out in two different tubes or gel posts, each tube or gel post has wild type reverse primer or mutant type, and both share a common forward primer. The bold base is the SNP sitting at the 3' end of the reverse primer. The Wild R or Mut R has a deliberately introduced mismatch at the penultimate base to increase the stringency of the ASPCR reaction. The product size is 213bp, and can be detected upon melt curve analysis.

Figure 2.2 Experimental strategy for ASPCR work with plasmid DNA template



The work in the conventional system was carried out to know the Tm (melting temperature value of the product) and the range of thermal cycling parameters. However, the thermal cycling conditions were re-optimized for the gel post environment, since it contains added reagents for the gel matrix formation and the technology uses Viriloc instruments (developed in our lab) for thermal cycling.

2.1.1 Materials and methods

2.1.1.1 Plasmid DNA, cloning and sequencing

Plasmids for the *4 SNP were obtained by cloning the fragment of genomic sequence containing the *4 allele. Genomic DNA samples were drawn from blood samples that were anonymously collected at University of Alberta Hospital from individuals not thought to have cancer. The blood samples were taken after approval from the University of Alberta and the Alberta Cancer Board and with informed consent. The DNA was isolated using QIAmp DNA Blood Mini Kit (Qiagen, #51106).

2.1.1.2 Preparing the samples for sequencing

For the PCR reaction 2µl of 25ng/µl purified genomic DNA sample was added to the 23µl of reaction mix containing 2.5µl of 10X PCR buffer, 1µl of 50mM $MgCl_2$, 0.5µl of each 10µM forward primer (F1 and R1) and 10µM of reverse primer (Integrated DNA Technologies, San Diego, CA) (Table 2.1), 0.5µl of Platinum *Taq* Polymerase (5U/µl), 0.5µl of 10µM dNTPs and 17.5 µl of distilled water. The PCR cycling parameters were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s and both annealing and extension at 72°C for 30s iii) final extension of 72°C for 120s. Post PCR

products were sent to Dr. Linda Pilarski's lab at the Cross Cancer Institute for sequencing of the genomic sample and cloning it to provide me with a pool of plasmid DNAs carrying known SNP alleles.

Table 2.1 Primers used in cloning the SNP site in plasmid DNA

S.No	Primer/Probe	Primer Sequence	Product Size
1	F1	5'aggcgcttctccgtctccac	216 bp
2	R1	5' agggaggcgatcacgttgct	

Forward primer F1 and reverse primer R1 were used to generate the PCR product containing *4 SNP site, and further short fragment flanking the SNP site in the PCR product was cloned to produce plasmid DNA carrying homozygous wild type or mutant SNP alleles.

2.1.1.3 Sequencing reaction

Sequencing of the DNA sample was performed on the Applied Biosystems Genetic Analyzer 3130XL. Once the sequencing was done for the **4* SNP (1846G>A), plasmids with the region of interest were cloned for homozygous wild type (G/G) and homozygous mutant allele (A/A). The cloned plasmids were received from Dr. Pilarski's lab. Whenever required, the

heterozygous condition was created by mixing equal amounts of homozygous wild and mutant plasmids.

2.1.2 ASPCR and MCA to detect SNP *4 in plasmid DNA

The PCR and MCA were carried out in a conventional and gel post system. The conventional system includes i) thermocycler for PCR ii) gel electrophoresis for sizing the PCR product iii) Lightcycler 2.0 (Roche) for profiling the MCA of the PCR product. The gel post system was capable of performing the PCR and MCA seamlessly without any manual intervention. The melting temperature of the product as defined by the Lightcycler was considered as standard, for comparing the results obtained conventionally with those obtained using the gel post system.

2.1.2.1 Plasmid controls in ASPCR reaction

In conventional and gel post systems, altogether seven different ASPCR reactions were carried out with different primer and plasmid DNA (match and mismatch) combinations. These included: two individual reactions of wild and mutant primer with the homozygous match pDNA, two individual reactions of wild and mutant primer with the heterozygous pDNA (match), and two individual reactions of wild and mutant primer with the homozygous mismatch pDNA and one no template control reaction with both the primers added within the same tube (conventional) or gel post but without plasmid DNA.

2.1.2.2 ASPCR/MCA on conventional system

The ASPCR was carried out in a thermocycler and the presence or absence of product and its size was confirmed by gel electrophoresis. In total eight individual PCR reactions were run.

i) Each ASPCR reaction 2µl of 25ng/µl plasmid DNA sample was added to the 23µl of reaction mix containing 2.5µl of 10X PCR buffer, 1µl of 50mM $MgCl_2$, 0.5µl of each 10µM forward primer (Common F) and 10µM of wild reverse (Wild R) or mutant reverse (Mut R) (Integrated DNA Technologies, San Diego, CA) (Table 2.2), 0.5µl of 10µM dNTPs, 0.5µl of Platinum *Taq* Polymerase (5U/µl) and 17.5 µl of distilled water. The ASPCR cycling parameters were i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s, annealing at 55°C for 20s and extension at 72°C for 20s iii) final extension of 72°C for 120s. Post ASPCR products were either immediately used to run gel electrophoresis or stored at -4°C for later use.

ii) Gel electrophoresis was done to confirm the size of the ASPCR product. The ASPCR products were run on 2% agarose for 55 minutes at 160V. The products were run along with the100bp DNA ladder as a reference ladder.

iii) 2µl of 10xLC Green Plus (Idaho Technology Inc., Salt Lake City, Utah) was added to the 18µl of each ASPCR plasmid product and melt curve analysis was performed on the Lightcycler 2.0 (Roche Applied Science, USA) from 50°C to 99°C.

Table 2.2 ASPCR primers used in plasmid DNA

S.No	Primer/Probe	Primer Sequence	Product Size
1	Common F	5'gcttctccgtctccaccttgc	213bp
2	Wild R	5' ggggcgaaaggggcgtat	
3	Mut R	5' ggggcgaaaggggcgtac	

Common forward primer (Common F) was shared between allele specific reverse primers (Wild R/ Mut R) and produces a 213 bp product. The red base in the reverse primer denotes a deliberate mismatch to increase the stringency of the reaction, and the blue base sits at the SNP base.

2.1.2.3 ASPCR and MCA on the gel post system

The melting temperature of the ASPCR product on the Lightcycler was considered as the 'gold standard' for comparison on the gel post system. The ASPCR and MCA were carried out seamlessly on the Viriloc instrument.

2.1.2.3.1 Making of gel posts on glass coverslip

Gel posts were printed on a coverslip from a glass mold that has a 9X9 array of wells. The mold was treated with Sigma Coat (Sigma, St. Louis, MO, cat no. SL2) and left to dry. The sigma coat prevents the mold from adhering to the gel post during polymerisation, and helps in easy detachment of the gel posts from the mold after polymerisation. PCR master mix contained azobis (2,2'-azobis(2-methyl-N-(2-hydroxyethyl) propionamide)) which triggers photo polymerisation when the mold was exposed to ultraviolet light. After loading the mold with the master mix, the wells were sealed by a glass coverslip (22X22 mm, Fisher, Fair Lawn, NJ, cat no. 12-54B) that was pre-treated in a mixture of 40mL of 95% ethanol, 1mL of 100% acetic acid (Fluka, Buchs, cat no. 45725) 100 μL of 3-(trimethoxysilyl)propyl methacrylate (Sigma, cat no. 440159), and , 8.9 mL of water for one hour followed by washing with isopropanol (2-propanol). Shortly after the polymerisation, the coverslip was delaminated from the mold. The coverslip carrying the gel posts was immersed in oil in a black aluminum pan and placed in the Viriloc instrument for PCR and melt curve analysis. The mold can be washed and reused for further casting gel posts on coverslips [1].

2.1.2.3.2 Viriloc instrument for thermal cycling

The Viriloc instrument is a prototype used to perform PCR thermocycling as the conventional thermocycler by employing a Motorola 68322 microprocessor that controls the heating element (Peltier element (XLT2398-01 L, Marlow Industries, Dallas, TX). A 405 nm laser diode was used as the fluorescence excitation source. A biconvex lens (KBX046, Newport, effective focal length 25.4 mm) and a 510 nm filter (BP510/50, Chroma Technology, Bellows Falls, VT) were placed between the camera and the gel posts to enhance the fluorescence capture. During each extension phase of the PCR cycle, the laser was turned on and the camera took a picture which was saved on the computer. During the MCA the laser was kept turned on and the images of the gel posts fluorescence were taken continuously at each degree from 55 to 95 °C.

2.1.2.3.3 Gel post reaction mix

The volume of each gel post was ~800nl. Initially a 100µl reaction mix was prepared that has mainly two types of reagents i) for gel post formation ii) for ASPCR reaction. A total of 100µl reaction mix contains: 2µl of 25ng/µl plasmid DNA, 20µl of 5X UB1 buffer, 10µl of 40% acrylamide, 10µl of 10X LC Green Plus (Idaho Technology Inc., Salt Lake City, Utah), 4µl of 50mM

$MgCl_2$, 3µl of *Taq*2000 (20 units/µL), 2µl of 10mM dNTPs, 2µl of 10µM each forward and reverse primer (wild or mutant primer pair) (similar primers that of the conventional system) (Integrated DNA Technologies, San Diego, CA) (Table 2.2), 2µl of 3% azobis (Sigma), 2µl of 1% BSA, 1µl of 10% TEMED (N,N,N′,N′-tetra-methylethylenediamine, Sigma, cat no. T7024) and 40 µl of distilled water.

2.1.2.3.4 PCR and MCA parameters

The ASPCR cycling parameters on the Viriloc were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 15s, annealing at 53°C for 30s and extension of 72°C for 30s iii) final extension of 72°C for 120s. Immediately after the ASPCR, MCA was set to start at 60°C, reaching 99°C at the end of the melt cycle.

2.1.2.3.5 PCR/MCA analysis of gel posts

The fluorescence variation in the gel posts was analysed using software imageJ software (National Institutes of Health, U.S.) that uses the Microarray Plug-in (Dr. Robert Dougherty, OptiNav Inc., Redmond, WA). The software uses Linear Regression of Efficiency (LRE) method for the PCR analysis and curve fitting. The melt curve analysis was achieved by the negative derivative of the fluorescence with respect to the programmed range for the

post-PCR melting of the DNA. Further, the linear offset for the melt curve is taken from the LRE based curve fitting [2].

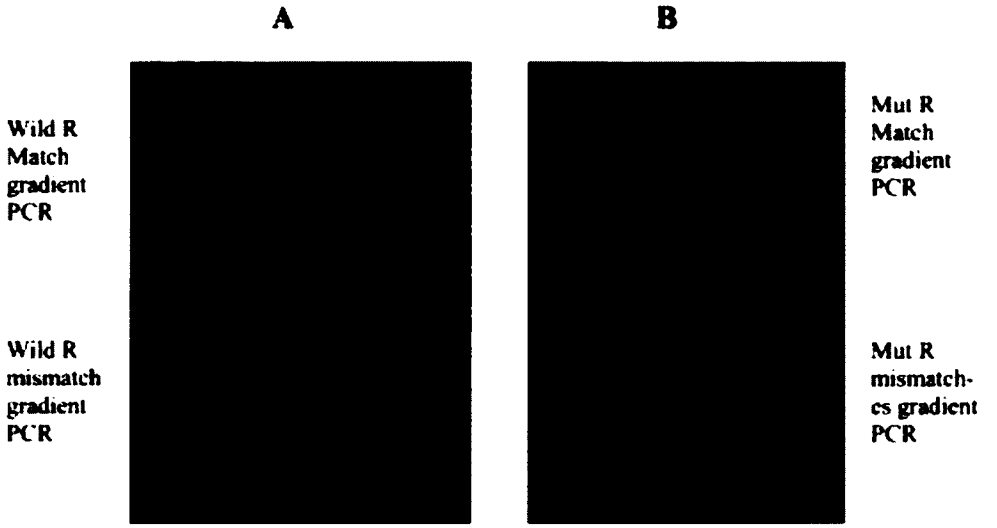
2.1.3 Results

The ASPCR strategy was successfully established using plasmid DNA in both conventional and gel post systems. The melting temperature of 93°C in gel post system was confirmed with the conventional system (Lightcycler 2.0). These primers were ready to be tested on genomic DNA. However, it was later realized that these primers could potentially bind to unwanted pseudogenes when using genomic DNA as templates. Hence, I proposed several methods to avoid pseudogenes and still detect the *CYP2D6*4* allele status.

2.1.3.1 Results in conventional system

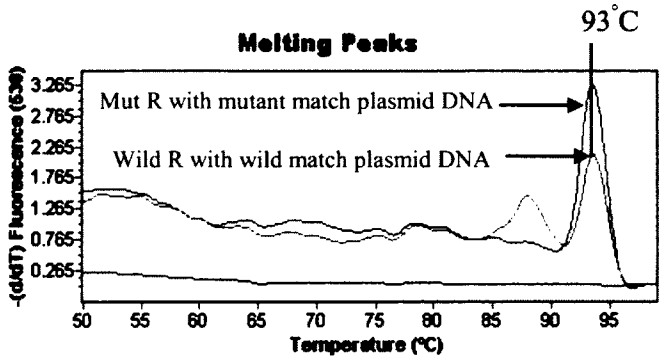
The allele specific wild and mutant primers were added separately in two different tubes along with the common forward primer. The primers were allele-specific for a wide range of temperature (65°C – 72°C) (Figure 2.3). However, a higher annealing temperature (70°C-72°C) was preferred as it avoided non-specific binding of the primers and formation of unintended short duplexes. Using the PCR product a melt curve analysis was performed on Lightcycler 2.0 at Provincial Labs, Edmonton (Figure 2.4).

Figure 2.3 Allele specific PCR work using plasmid DNA in conventional system (PCR on thermocycler)



Results of gradient ASPCR with wild type and mutant primer using plasmid DNA. A. wild type allele specific reverse primer (Wild R) amplified when matched with plasmid (wild type pDNA, top panel), and primer did not amplify in case of mismatch (mutant pDNA, bottom panel) B. In the upper panel of the gel the mutant reverse primer matches the SNP allele (A), and did not extend due to the mismatch (bottom panel), hence no product was formed.

Figure 2.4 Allele specific PCR work using plasmid DNA in conventional system (melt curve analysis on Lightcycler)

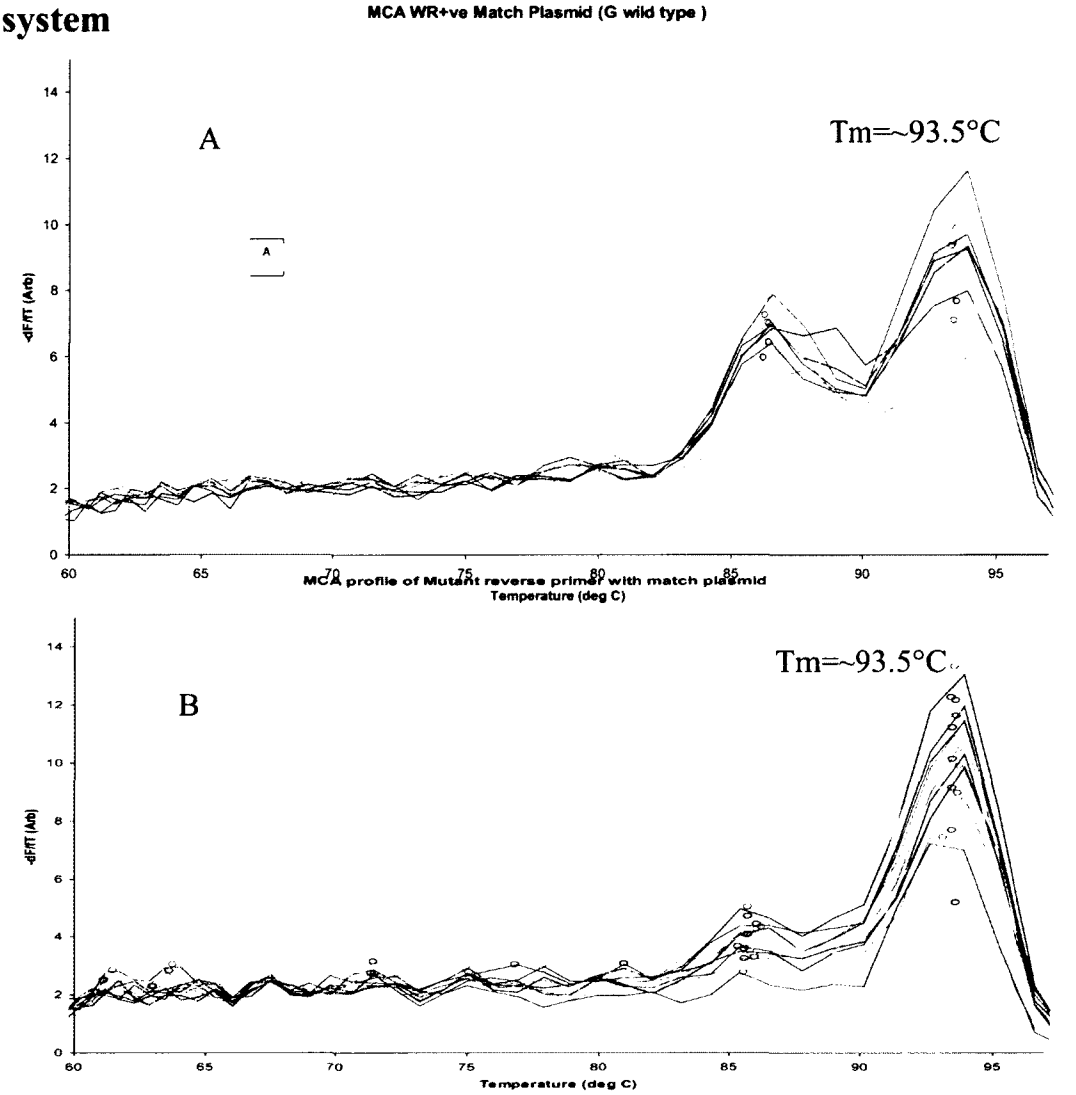


Melt curve analysis of thermocycler PCR product of wild type reverse and mutant type reverse allele specific primers to check the melting temperature of the product. The melting temperature of the PCR product was found to be 93°C.

2.1.3.2 Results in gel post system

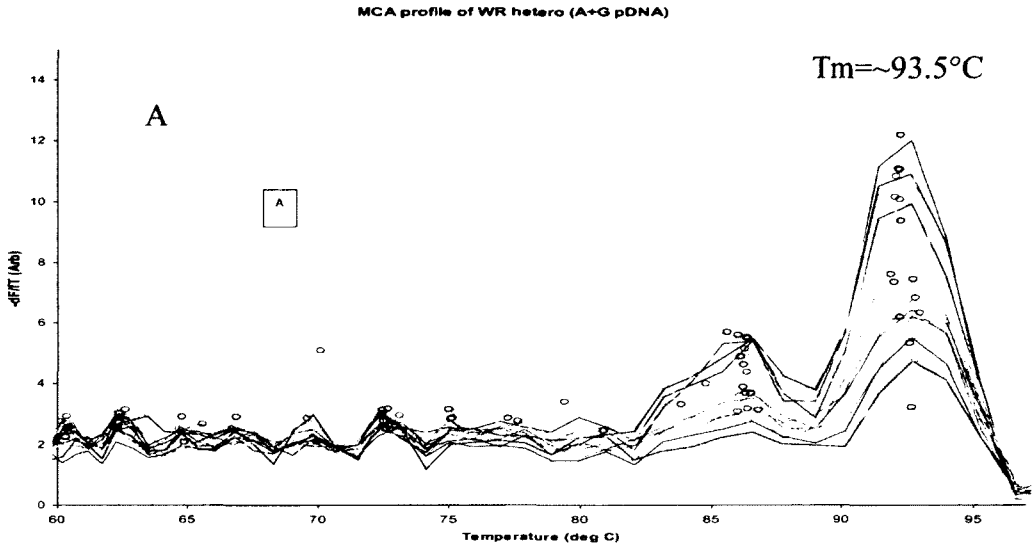
The optimized PCR thermal cycling conditions required adjustment for the gel post system due to 1) use of a gel matrix system instead of liquid PCR, 2) use of the Viriloc instrument instead of a thermocycler for PCR cycling and 3) addition of reagents to create the gel posts. Optimizing the conditions in conventional system was still important as this provides the range of temperatures for initial testing of PCR in a gel post system, and the melting temperature (Tm) of the PCR product as the standard to determine the Tm of the product in the gel post system. The Tm in the gel post system was found to be +0.5°C, which lies within the range of the Viriloc Tm error zone (Tm±1°C). Allele specific PCR results in the plasmids are shown for primer-pDNA match (Figure 2.5 and Figure 2.6) and primer-pDNA mismatch and no-template-control (Figure 2.7).

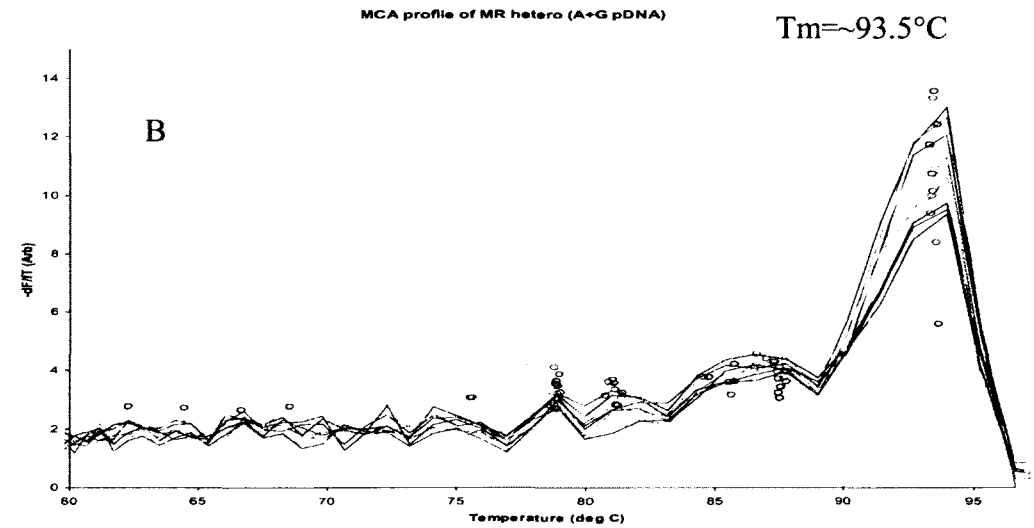
Figure 2.5 Melt curve analysis of allele specific PCR (primer-homozygous DNA match) product using plasmid DNA in gel posts system



ASPCR was carried out for the wild type primer and mutant primer separately in distinct posts, each with wild type pDNA or mutant pDNA respectively. The curve was drawn for the negative value of the raw fluorescence value over time against temperature. A) The wild type primer matches with the homozygous wild type allele (wt/wt) B) The mutant primer matches with the mutant allele (mt/mt). PCR products were the same in both reactions and they yielded the same melt curve. The product melts at 93.5°C. The extra bump at around 86°C prior to the real product melt curve might be a primer dimer, which was also seen in the conventional setting

Figure 2.6 Melt curve analysis of allele specific PCR (primer-heterozygous DNA match) product using plasmid DNA in gel posts system

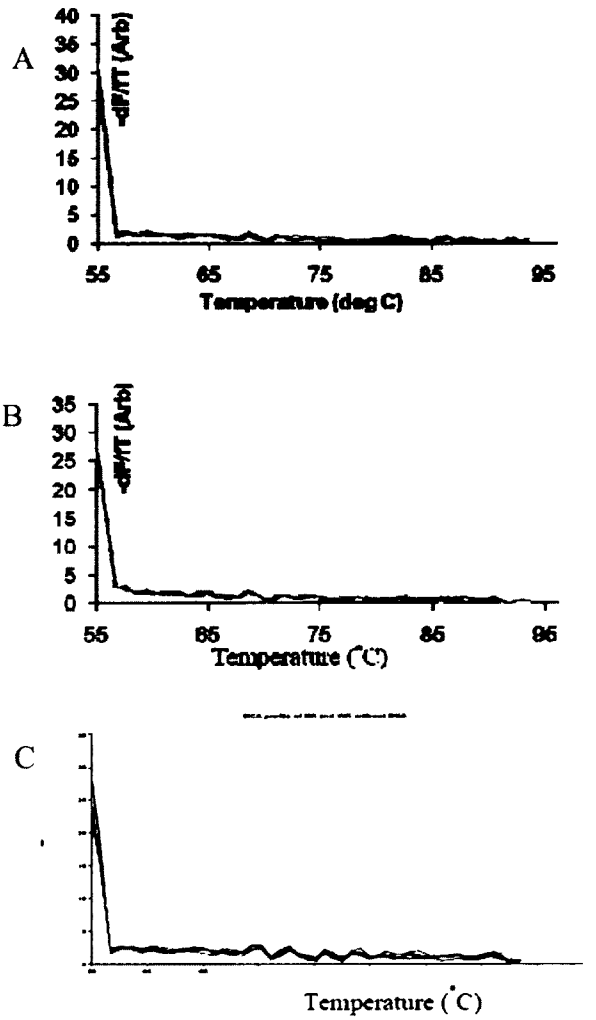




ASPCR was carried out for the wild type primer and mutant primer separately in distinct posts, each with heterozygous plasmid DNA, created by mixing equal amounts of homozygous wild pDNA and mutant pDNA. A) The wild type primer matches with the heterozygous sample (wt/mt) B) The mutant primer matches with the heterozygous sample (wt/mt). The curve was drawn for the negative value of the raw fluorescence value over time against temperature. The PCR and MCA was optimized to obtain relative

PCR product and peak strength, since heterozygous DNA halves the amount of match DNA available to the respective wild or mutant type allele specific primer.

Figure 2.7 Melt curve analysis of allele specific PCR (primer-DNA mismatch) product using plasmid DNA in the gel posts system



ASPCR was carried out for the wild type primer and mutant primer separately in distinct posts, each with heterozygous plasmid DNA, created by mixing equal amounts of homozygous wild pDNA and mutant pDNA. The curve was drawn for the negative value of the raw fluorescence value over time against temperature. A) Wild type primer with mutant pDNA B) mutant primer with the wild type pDNA C) wild and mutant primer together with the non-template-control (NTC). In all the cases no PCR product was obtained due to the mismatch between the reverse primer and the specific template.

2.1.4 Conclusion

The ASPCR strategy was successfully developed using plasmid DNA. I could not eliminate the primer dimer peak (at ~86°C) in the melt curve. However, recognition of the match and mismatch SNP allele by primers has been achieved both conventionally and using the in-gel method. I abandoned these primers for use with a genomic DNA template, due to the limitations and challenges posed by the pseudogenes. The details of detecting SNPs in genomic DNA are discussed in the next chapter.

References

[1] A. Atrazhev, D. P. Manage, A. J. Stickel, H. J. Crabtree, L. M. Pilarski and J. P. Acker, "In-Gel technology for PCR genotyping and pathogen detection," *Anal. Chem.*, vol. 82, pp. 8079-8087, 2010.

[2] R. G. Rutledge and D. Stewart, "Assessing the performance capabilities of LRE-based assays for absolute quantitative real-time PCR," *PLoS One,* vol. 5, pp. e9731, 2010.

Chapter 3. Characterization of genomic DNA

Before developing ASPCR for genomic DNA, firstly it was important to 1) screen the samples for deletion or duplication of the active gene, 2) characterize the population for *4 allele condition by sequencing the relevant region of CYP2D6 for a set of 28 human blood samples to provide template DNA preparations with known SNP genotypes and 3) develop an ACPCR test to detect the presence of *4 variant on the active gene and exclude the pseudogenes, using the set of samples with a known genotype.

3.1 Step one: To screen genomic samples for deletion, duplication, and single gene carriers

For this purpose, I used six published primers (3 pairs). Each primer pair is used for detecting carriers of single active CYP2D6 gene, duplication or deletion event. Once the samples are grouped, they can be used as controls carrying known SNPs, deletions and/or duplications, to verify the accuracy of the PCR product that is amplified.

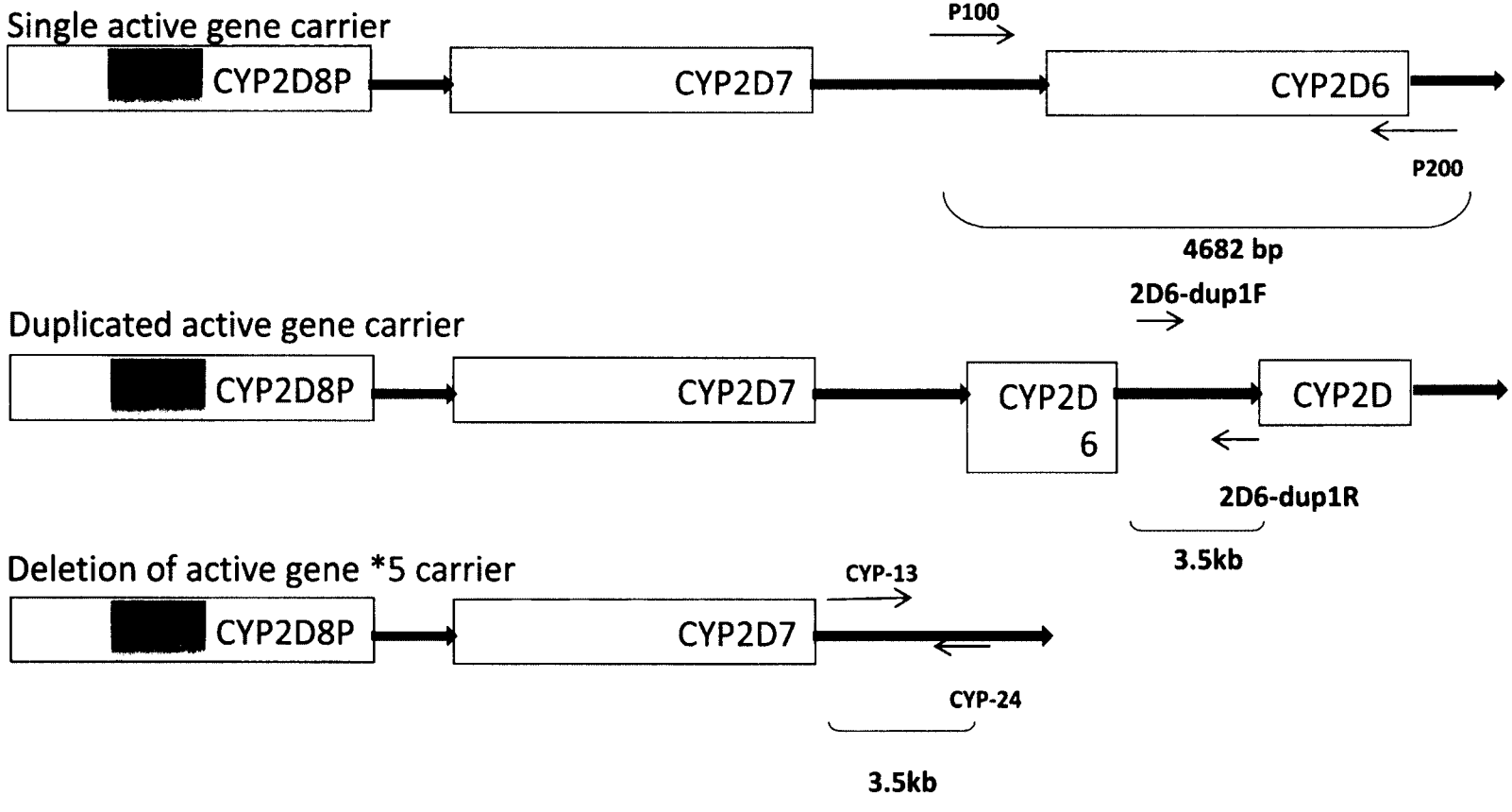
3.1.1 Strategy

The first primer pair (P100 and P200) is used to amplify a single CYP2D6 gene product carrying all the nine exons. The second primer pair (dupF-dupR) binds to the CYP2D6 duplication specific sites, and it is used to detect active gene duplication. The third primer pair (CYP13-CYP24) is used to detect complete

deletion of the gene. The first primer pair makes a 4682 bp PCR product if a single active gene is present. In case of duplication or deletion, the other two primer pairs make a 3.5kb PCR product. In this step, the distribution of *5 allele (deletion) in the samples is identified. The primers are taken from different published articles [1-3].

Additional sets of primer pairs were used to detect the *4 SNP. These primers bind away from the SNP site and generate a 201 bp product. The PCR product can be sequenced to reveal the SNP base (wild or mutant type) (Figure 3.1).

Figure 3.1 Screening of the population for single gene, deletion and duplication of the *CYP2D6* gene



The primer pair P100-P200 is used to amplify a single gene product for further characterization. The 2D6-dup1F and 2D6-dup1R primer pair detects the duplication event. The CYP13 and CYP24 primer pair is used to detect the deletion of the active gene. The single gene PCR product size is 4682bp, whereas the deletion and duplication PCR products sizes are of 3.5Kb.

3.1.2 Materials and methods

3.1.2.1 DNA samples

Genomic DNA was purified from 28 blood samples taken at the Cross Cancer Institute used here as anonymous samples for developing a set of controls with a known genotype. The samples were taken after approval from University of Alberta and the Alberta Cancer Board, and with informed consent. The DNA was isolated using the QIAmp DNA Blood Mini Kit (Qiagen, #51106). The purified genomic DNA samples were further used to amplify the single gene and characterize deletion and/or duplication events and allele *4 status, and ultimately to serve as known controls for developing the allele specific PCR reaction on gel posts.

3.1.2.2 Long range PCR for screening the population

Three individual long range PCR were performed to identify the samples for 1) single gene carriers 2) duplication carrier and 3) complete deletion of the gene respectively as per the published protocol [4]. All three independent reaction mixes have similar PCR recipes except for the primer pairs.

i) PCR protocol to identify single gene carriers: 25µl PCR mix is prepared for each genomic sample. PCR mix contained 2.5µl of 100ng/µl purified genomic DNA sample, 1.25 µl of each 10µM

primer P100, P200 (Integrated DNA Technologies, San Diego, CA) (Table 3.1) 10µl of reaction buffer (Invitrogen, CA, USA), 0.75µl of 50mM $MgCl_2$, 0.5µl of 10mM dNTPs, 2µl Elongase Enzyme Mix (Invitrogen, CA, USA) and 6.75 µl of distilled water. The PCR cycling parameters were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s, annealing and extension at 70°C for 6min and iii) final extension of 70°C for 10min.

ii) PCR protocol to identify active gene deletion and duplication carriers: for each sample two individual reactions were carried out to detect deletion and duplication. The PCR reaction mix is the same as above except for the primer pairs. To detect the deletion, 0.5µl of primers CYP-13 and CYP-24, and to detect duplication, 0.5µl of 2D6-duplF and 2D6-duplR, were added separately

iii) Gel electrophoresis was done to confirm the size of the long range PCR product. The PCR products were run on 1% agarose for 1 hour at 100V against a 10kb ladder.

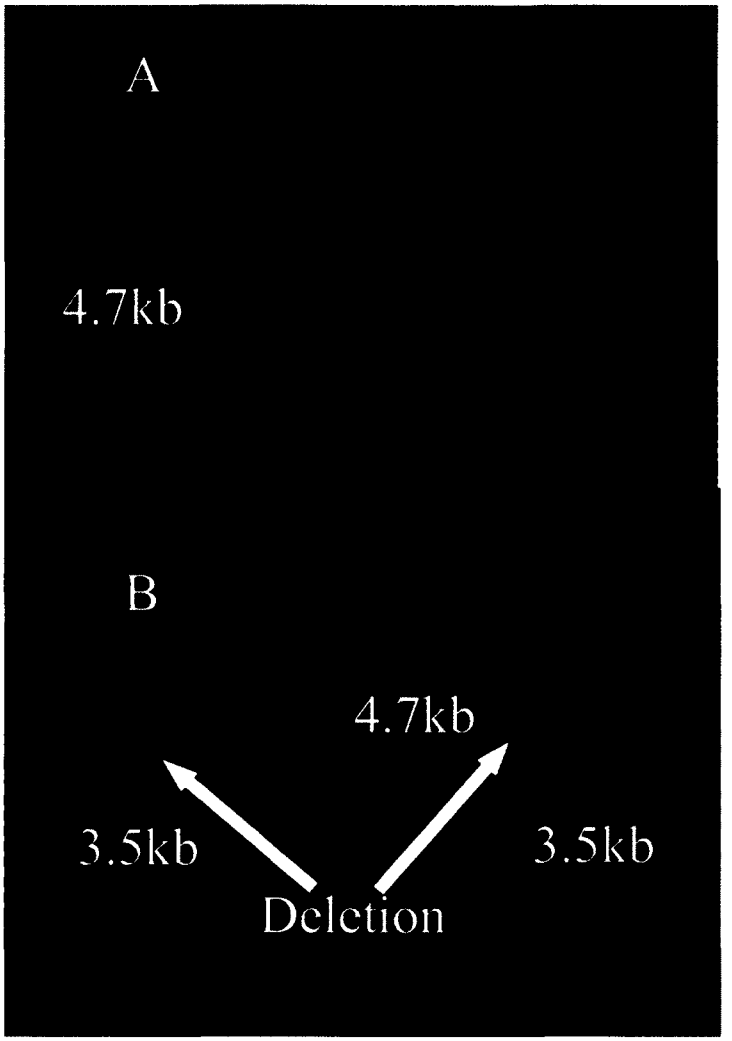
Table 3.1 Long range PCR primers for screening the population singe gene, deletion and duplication. These primers are taken from published work [1-3]

S.No	Primer	Primer Sequence	Product Size
Primer pair for characterising the single gene			
1	P100	5' ggcctaccctgggtaagggcctggagcagga	~4.7kb
2	P200	5' ctcagcctcaacgtacccctgtctcaaatgcg	
Primer pair for characterising gene deletion			
3	CYP-13	5'accgggcacctgtactcctca	3.5kb
4	CYP-24	5'gcatgagctaaggcacccagac	
Primer pair for characterising gene duplication			
5	2D6-DuplF	5'cctgggaaggccccatggaag	3.5kb
6	2D6-DuplR	5'cagttacggcagtggtcagct	

3.1.3 Results

The single gene carrier samples produced PCR products of ~4.7kb in size, whereas the deletion and duplication carriers produced a 3.5kb of PCR product. Only 2/28 individuals carried deletions, with the rest being single gene carriers (Figure 3.2). There were no duplications found in the population as the frequency of duplication is very low in the Caucasian population (Table 3.2). The screened genomic samples were further characterized for the SNP status and developing ASPCR.

Figure 3.2 Gel electrophoresis of the long range PCR for population screening



26/28 individuals were single gene carriers and produced a 4.7kb PCR product, whereas 2/28 individuals had complete deletion of the entire CYP2D6 gene and produced a 3.5kb PCR product. Samples in the upper electrophoresis gel (A) has all the individuals showing a 4.7kb and the lower electrophoretic gel (B) shows all individuals carrying single gene except 2 deletions producing a 3.5kb PCR product.

Table 3.2 Population screening for single gene, deletion and duplication status

Number of genomic samples	Single gene carriers	Deletion carriers	Duplication carriers
28	26	2	0

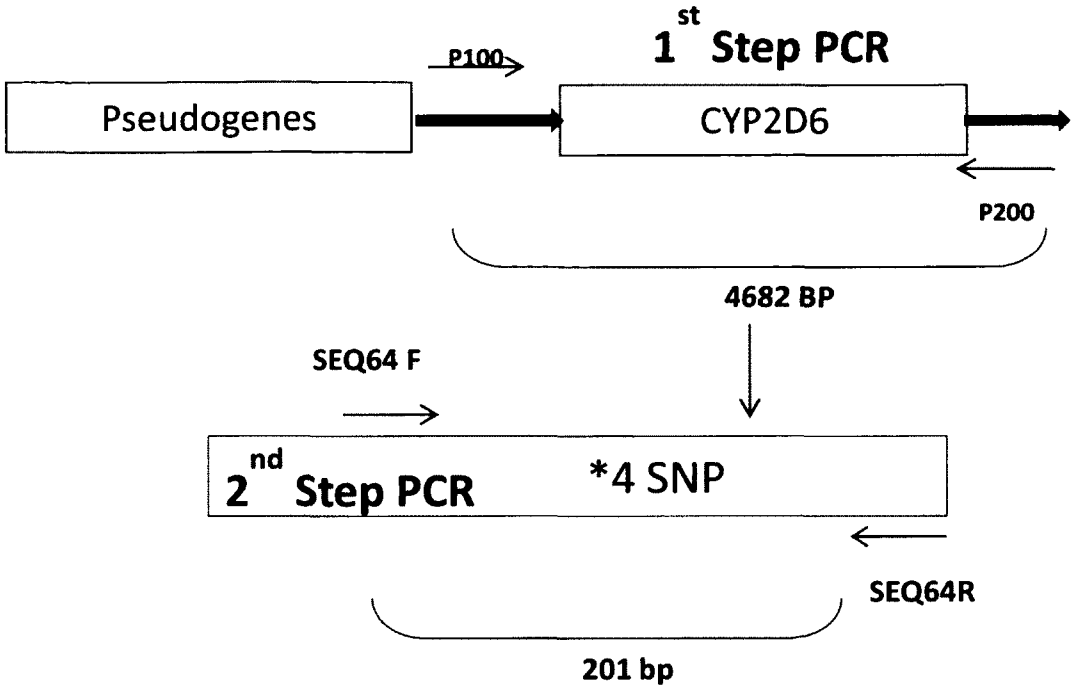
3.2 Step two: To characterize genomic samples for *4 genotype (wt/wt, wt/mt, mt/mt)

The control genomic samples were characterized for the *4 genotype so that they can serve as known templates when developing ASPCR test for wild type, mutant type and heterozygous DNA.

3.2.1 Strategy

SNP genotyping was achieved in two separate PCR steps (Figure 3.3). The first PCR used the primers P100 and P200 to amplify only CYP2D6 gene. This primer pair prevents the contamination of the PCR product by amplification of CYP2D7 or CYP2D8P pseudogenes. The PCR product provides the DNA template for the subsequent PCR reaction, where newly designed primers SEQ64F-SEQ64R were used to make a PCR product of size 201bp (Figure 3.4), flanking the *4 SNP site. The PCR product was sent to Dr. Pilarski's lab for conventional sequencing for *4 alleles.

Figure 3.3 Screening the samples for the *4 allelic condition



The first step PCR excludes the amplification of the pseudogene, and the subsequent PCR amplifies the SNP region. Once the samples are sequenced, they can be used as known controls for developing ASPCR test to detect the status of *4 SNP.

Figure 3.4 2nd step PCR primer that produces a 201bp carrying the SNP site. The SNP site (G/A) is shown in bold

CGCAACTTGGGCCTGGGCAAGAAGTCGCTGGAGCAGTGGGTGA

CCGAGGAGGCCGCCTGCCTTTGTGCCGCCTTCGCCAACCACTCCGGTG

GGTGATGGGCAGAAGGGCACAAAGCGGGAACTGGGAAGGCGGGGGA

CGGGGAAGGCGACCCCTTACCCGCATCTCCCACCCCCAG/AGACGCCC

CTTTCGCCCCAACGGTCT

3.2.2 Materials and methods

The PCR is accomplished in two steps i) The first PCR is a long range PCR that produces a 4.7kb product to make sure that the PCR product is from the active gene and excludes any contamination of pseudogenes, ii) The 2nd PCR reaction uses the PCR product from the first reaction as a template to amplify a small region flanking the SNP. Later the PCR products were sent to Dr. Pilarski's lab for sequencing.

3.2.2.1 Conventional PCR

i) Individual reactions were carried out for each genomic sample. The PCR mix contained 2.5µl of 100ng/µl purified genomic DNA sample added to the 22.5µl of the master mix containing 1.25 µl of each 10µM primer P100, P200 (Integrated DNA Technologies, San Diego, CA) (Table 3.1), 10µl of reaction buffer (Invitrogen, CA, USA), 0.75µl of 50mM $MgCl_2$, 0.5µl of 10mM dNTPs, 2µl Elongase Enzyme Mix (Invitrogen, CA, USA) and 6.75 µl of distilled water. The PCR cycling parameters were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s, annealing and extension at 70°C for 6min and iii) final extension of 70°C for 10min.

ii) Gel electrophoresis was done to confirm the single gene of size ~4.7kb. The PCR products were run on 1% Agarose for 1 hour at 100V against a 10kb ladder.

iii) 2µl of the first PCR product is added to the 23µl of master mix containing 2.5µl of 10X PCR buffer, 1µl of 50mM $MgCl_2$, 0.5µl of each 10µM forward primer (SEQ64F) and 10µM of reverse primer (SEQ64R) (Integrated DNA Technologies, San Diego, CA) (Table 3.3), 0.5µl of 10µM dNTPs, 0.5µl of Platinum *Taq* Polymerase (5U/µl) and 17.5 µl of distilled water. The PCR cycling parameters were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s, annealing at 55°C for 20s and extension at 72°C for 20s iii) final extension of 72°C for 120s.

iv) The PCR products are confirmed by running gel electrophoresis on 2% Agarose for 55min at 160V against 1kb ladder.

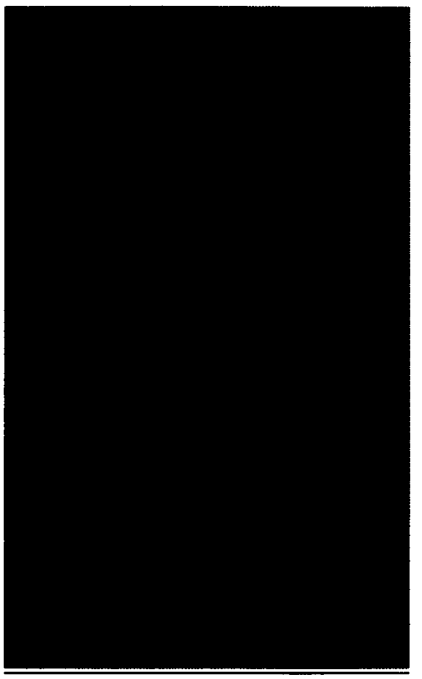
Table 3.3 Primer for 2nd step PCR

S.No	Primer	Primer Sequence	Product Size
1	SEQ64F	5'cgcaacttgggcctgggcaa	201bp
2	SEQ64R	5'agaccgttggggcgaaaggg	

3.2.3 Results

Out of 28 samples 2 samples had deletions and did not show any PCR product in the 2 step PCR. All the other samples showed a 201bp product (Figure 3.5). The distribution of the SNP in the population is shown in Table 3.4. The SNP distribution in the set of control samples is as follows: 14/28 individuals were heterozygous, 4/28 individuals were homozygous mutant and 8/28 individuals were homozygous wild type.

Figure 3.5 2nd step PCR product flanking to the SNP *4 allele



Electrophoresis of the amplicon from four of the 26 genomic samples shows the 201bp product produced in the 2nd step PCR. The 2nd step PCR uses the long range PCR product as template to produce a 201bp product flanking to the SNP of allele *4. Electrophoresis of the 4/26 genomic samples is shown here against 100bp DNA ladder.

Table 3.4 Population screening for the CYP2D6*4 SNP rs 3892097 status

Number of genomic samples	Gene deletion	Homozygous wild type (G/G)	Homozygous mutant type (A/A)	Heterozygous (G/A)
28	2	8	4	14

3.2.4 Conclusion

Pseudogenes are highly similar to the active gene; hence it was important to screen the population for single gene and deletion carriers in order to identify any false negatives- defined here as samples that appear to be wild type, lacking the *4 allele, because the pseudogene was inappropriately amplified. Use of these known controls will determine the extent to which false negatives arising from pseudogene amplification are occurring in the ASPCR test. This would be impossible without the set of genotyped control samples to validate ASPCR test results. Out of the 28 genomic samples, 2 samples had deletions of the entire CYP2D6 gene. These samples were removed from the panel for developing the *4 detection strategy, because they fail to yield a PCR product with either set of allele specific primers.

The primers used to screen the population for single gene amplification were taken from the published work, and they were also useful in screening the population for the SNP status so that the samples can serve as known controls.

These primers generated a long PCR product of size 4.7kb which was then used as the template for a subsequent nested PCR to shorten the product size to 210kb for sequencing purposes. The sequencing of the PCR product revealed the SNP status of the control group, providing a set of controls with known deletions and SNP status. The screening information was very useful in developing the actual SNP detection test.

References

[1] C. Sachse, J. Brockmöller, S. Bauer and I. Roots, "Cytochrome P450 2D6 variants in a Caucasian population: Allele frequencies and phenotypic consequences," *Am. J. Hum. Genet.*, vol. 60, pp. 284-295, 1997.

[2] V. M. Steen, O. A. Andreassen, A. K. Daly, T. Tefre, A. Brresen, J. R. Idle and A. Gulbrandsen, "Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology," *Pharmacogenetics,* vol. 5, pp. 215-223, 1995.

[3] E. Lundqvist, I. Johansson and M. Ingelman-Sundberg, "Genetic mechanisms for duplication and multi-duplication of the human CYP2D6 gene and methods for detection of duplicated CYP2D6 genes," *Gene,* vol. 226, pp. 327-338, 1999.

[4] R. L. Roberts and M. A. Kennedy, "Rapid detection of common cytochrome P450 2D6 alleles in Caucasians," *Clin. Chim. Acta,* vol. 366, pp. 348-351, 2006.

Chapter 4. ASPCR with genomic DNA

4.1 Application of the plasmid ASPCR test to genomic DNA

The use of ASPCR for detecting CYP2D6 alleles was verified using cloned plasmids carrying the SNP along with the flanking sequence by designing the primers specific to either mutant or wild type allele (Chapter 2). When the same primers were used with purified genomic DNA as template, they failed to amplify the target region. Several bases were then added to the allele specific primers to increase the specificity of the primers for the target sequence during the PCR. However, even with added bases the test failed; the primers were able to amplify a product but failed to show the allele specificity in the presence of mismatch target sequence in the active gene.

The essential reason for the failure of primers to be allele specific when used with genomic DNA template is the presence of pseudogenes, which are highly homologous with the active gene, and will thus contaminate amplicons derived from the active gene, making interpretation of the test impossible. The presence of pseudogenes forced the development of ASPCR strategies that exclude amplification of pseudogenes, used here to develop an accurate test for SNPs on the active gene. Most of those strategies were abandoned due to specificity issues with the ASPCR primers, where primers amplify the correct product size, but could not discriminate between the wild and mutant alleles. For

further details on these abandoned approaches using various primer designs, please refer to the Appendix.

4.2 Challenges posed by pseudogenes for detection of the *4 allele

The pseudogenes carry a sequence fully complementary to the designed allele specific primers. The primers can amplify a pseudogene or/and the active gene, which is predicted to give false test results. Therefore, I analyzed the unique bases in the two pseudogene sequences that are variant to the CYP2D6 active gene (variation sites). Considering these variation sites as the 3' binding site for the common forward primer of the ASPCR primer pair, I designed several primers and tested them with genomic DNA. Unexpectedly, they did not show allele specificity. The reason might be that the variation sites in the pseudogenes are not common in the population. Therefore, I had to consider only the mutant base that actually gave rise to the evolution of pseudogenes as the potential 3' binding site for the common forward primer of the ASPCR primer pair. The mutant base is the prime cause of inactivation of the active gene (*CYP2D6*), and it is present in the pseudogenes (*CYP2D7 and CYP2D8PP*) of every individual.

In conclusion, the ASPCR has been designed with two different strategies to exclude the pseudogenes *CYP2D8P* and *CYP2D7* while amplifying the correct target gene *CYP2D6.* 1) Considering variation sites in pseudogenes as a 3'mismatch for common forward primer; this approach was abandoned. 2) Considering mutation sites instead of variation sites as 3'mismatch for forward

primer; this second approach gave allele specific results. However, it produced a secondary product along with the actual product.

4.3 Evolution of strategies to genotype *CYP2D6*4* and exclude pseudogene amplification

To determine the tamoxifen metabolism status, I proposed three different methods as test strategies for amplifying the region carrying the *4 SNP site and detecting the SNP. All three methods exclude the pseudogene amplification, and target only the active gene. They are 1) Allele specific PCR approach (ASPCR) 2) Asymmetric PCR with Locked Nucleic Acid (LNA) probe 3) Tetra-Allele Refractory Mutation System PCR (tetra-ARMS) test. The LNA based method proved to be less useful; hence it was added to the Appendix. The ASPCR strategy had distinct limitations when it comes to its their implementation on the gel post system, as discussed in the following text. However, the tetra-ARMS PCR test was reliable in a conventional system, and appears to overcome the limitations of the first two tests. Here I discuss the ASPCR approach, and then describe the extension of this strategy to the new tetra-ARMS PCR test. As the method of ASPCR led to the concept of modifying it to a tetra-ARMS strategy, the methods and results of these two methods are discussed in this chapter.

All the tests were developed in a conventional system. The amplification of the target active gene was achieved using a thermocycler, followed by confirming the correct size of the PCR product by gel electrophoresis. The MCA was performed on the Roche Lightcycler 2.0 to determine the product Tm. On the

gel post system, PCR and MCA were performed in a seamless fashion on Viriloc instruments, developed in our lab.

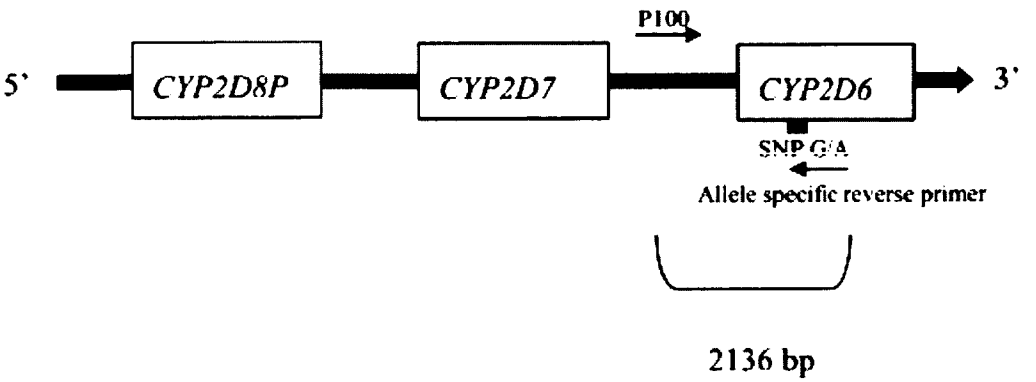
4.3.1 Allele specific PCR strategies

4.3.1.1 ASPCR strategy I

I proposed a ASPCR strategy (strategy I) as seen in Figure 4.1 to accomplish two tasks in a single PCR run 1) exclusion of pseudogenes during PCR 2) detection of SNP alleles by allele specific PCR. Theoretically this can be achieved by employing a common forward primer that binds to the *CYP2D6* specific intronic region. For this purpose a published primer was employed coupled with an allele specific reverse primer. The PCR product size was 2136 bp. I learned from our lab group work that it is difficult to achieve a PCR with 2136 bp size in the gel post environment, which is our ultimate goal. Hence, to reduce the size further, I went to the next allele specific strategy (strategy II) as shown in Figure 4.3. Please refer to the section 4.3.1.2.

A common forward primer (P100) was taken from a published work [1] that binds to the sequence upstream of the SNP base which is not homologous to the pseudogene, along with a reverse allele specific wild (WR42G) or mutant primer (MR4+9) combination (Table 4.1). The primers in either case (wt/wt, mt/mt or heterozygous) should make a product of 2136 bp in length.

Figure 4.1 ASPCR Strategy I



The common forward primer, taken from a published article [1] binds to the *CYP2D6* specific intronic region, and can be coupled with the allele specific reverse primer. The size of the PCR product is 2136 bp.

4.3.1.1.1 Materials and methods

i) ASPCR protocol: 25µl of PCR mix was prepared with 2.5µl of 100ng/µl purified genomic DNA sample, 1.25 µl of each 10µM primer P100 paired with either WR42G or MR4+9 in two different reactions for each genomic sample (Integrated DNA Technologies, San Diego, CA) (Table 4.1), 5 µl of each buffer A and buffer B (Invitrogen, CA, USA), 0.75µl of 50mM $MgCl_2$, 0.5µl of 10mM dNTPs, 2µl Elongase Enzyme Mix (Invitrogen, CA, USA) and 6.75 µl of distilled water. The PCR cycling parameters were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s,

annealing and extension at 70°C for 6min and iii) final extension of 70°C for 10min.

ii) Gel electrophoresis was done to confirm the size of the long range PCR product. The PCR products were run on 1% agarose gel for 1 hour at 100V against a 10kb ladder.

Table 4.1 Primers for ASPCR strategy I

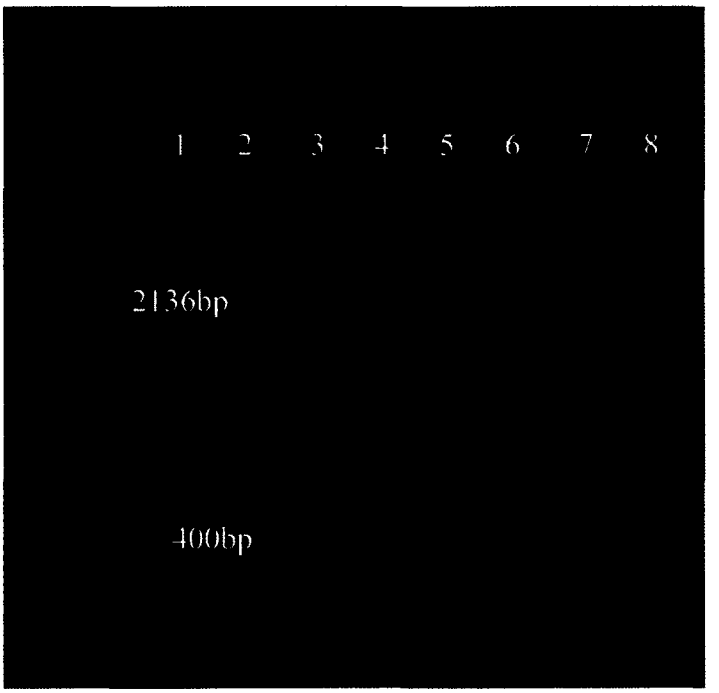
S.No	Primer	Primer Sequence	Product Size
Forward primer that excludes pseudogenes and amplifies only the active gene			
1	P100	5'ggcctaccctgggtaagggcctggagcagga	~2.1kbp
Reverse allele specific primers			
2	WR42G	5' cgttggggcgaaaggggcgtgc	
3	MR4+9	5'ccaagagaccgttggggcgaaaggggcgtgt	

P100 is the forward primer that excludes pseudogenes and amplifies active gene only and WR42G and MR4+9 work as reverse allele specific wild and mutant primers respectively. The red base in the reverse primer denotes a deliberate mismatch to increase the stringency of the reaction, and the blue base sits at the SNP base.

4.3.1.1.2 Results

The gel electrophoresis (Figure 4.2) shows the 2136 bp ASPCR product size of 2136bp. The reaction was successful for all 8/8 wild type (wt/wt) and 4/4 mutant (mt/mt) samples. However, this strategy detected only 12/14 heterozygous (wt/mt) samples. This might be due to the ASPCR being performed with such a large PCR product. I also found some secondary product formation along with the actual product in all of the reactions. This could be due to the stringency of allele specific reaction during the long range PCR reaction where primers were able to amplify only a portion of DNA over the course of the reaction. However, the major challenge was to reduce the ASPCR product size, which is discussed in the strategy II.

Figure 4.2 Gel electrophoresis of ASPCR product from strategy I with genomic DNA



Lanes 1-4 are primer-DNA matched product of size 2136bp (lane 1=wild type primer with wt/wt sample, lane 2=mutant primer with mt/mt sample, lane 3= wild type primer with heterozygous sample, lane 4= mutant primer with heterozygous sample). There was a secondary unintended product amplification of size ~400bp. To reduce the size of the ASPCR product strategy II was proposed. Lanes 5 and 6 are the primer-DNA mismatches with wild type primer and mutant primer respectively. Lanes 7 and 8 are the no-template control with the wild type and mutant primer respectively.

4.3.1.2 ASPCR strategy II

In the new strategy (Figure 4.3) the common forward primer was designed to bind at its 3' end to the mutation site of the pseudogenes, so that at its 3' end the primer is only complementary to *CYP2D6*, and mismatches to the

pseudogenes. This helps in preventing its extension in the pseudogenes. In this way, the product size is further reduced to 1752 base pairs.

Strategy II primers include CYP1, CYP2 and CYP3; they bind at the same location, except that they differ in the placement of a deliberate mismatch near to 3' end of the primer. CYP3 outperformed the other two primers by producing a strong PCR product. Hence, I will discuss here my work in relation to the forward primer CYP3. CYP3 works as a common forward primer and pairs with allele specific reverse primers WR42G and MR4+9 (Table 4.2).

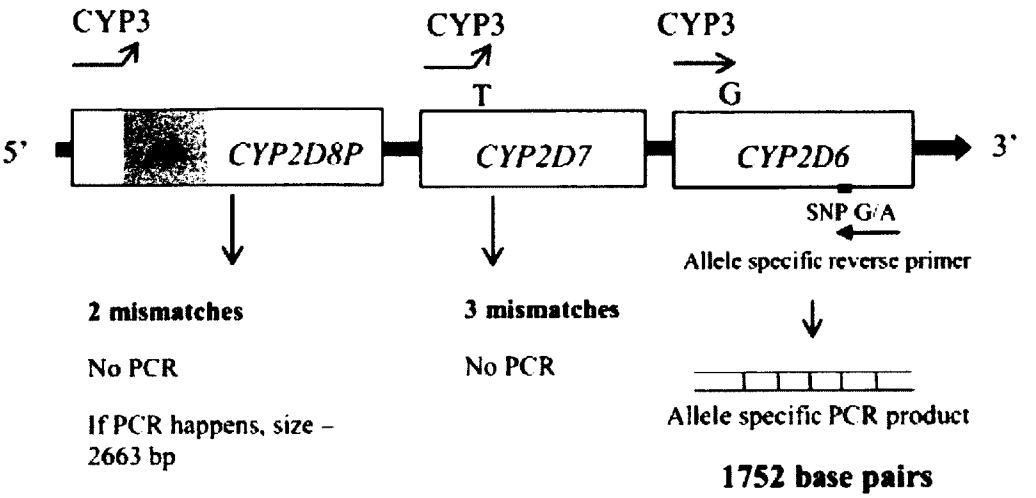
The 3' end of CYP3 is complementary to CYP2D6 only, since CYP2D7 bears a mutation (insertion of T) and CYP2D8P has multiple variations in the sequence at the 3' end of the primer. In total, it contains 3 mismatches with the pseudogene *CYP2D7* and 2 mismatches with the *CYP2D8P*. The specificity of the primer to only CYP2D6 was confirmed by sizing the PCR product on gel electrophoresis. The primer can produce the product size of 1752 bp only when extended within the active gene, which was confirmed with gel electrophoresis results. If the primers bind to the unintended pseudogene CYP2D8P, the product size becomes 2663bp due to the insertion of Alu sequence within the PCR target region in the first intron. The size of the ASPCR cannot be reduced any further, since there is only one mutation (transversion of G to T) found in the pseudogene that can be employed as a tool to exclude pseudogenes by the inability of a primer to bind the mutant base of pseudogene.

Table 4.2 Primers for ASPCR strategy II

S.No	Primer	Primer Sequence	Product Size
Forward primer that reduces the size of ASPCR product but still excludes pseudogenes from the reaction and amplifies only the active gene			
1	CYP3	5'actgcccgggctgggcaaccagc	~1.7kb
Reverse allele specific primers			
2	WR42G	5' cgttggggcgaaaggggcgtgc	
3	MR4+9	5'ccaagagaccgttggggcgaaaggggcgtgt	

CYP3 is the forward primer sequence that matches only with the active gene sequence at its 3' end and amplifies active gene only. WR42G and MR4+9 work as reverse allele specific wild and mutant primers respectively. The red base in the reverse primer denotes a deliberate mismatch to increase the stringency of the reaction, and the blue base sits at the SNP base.

Figure 4.3 ASPCR strategy II



The forward primers were designed to exclude the amplification of pseudogenes, considering the binding site of the 3' end of the primer is located at the mutation bearing site in the pseudogene *CYP2D7*(base G in the active gene undergo transversion to base T in pseudogene). The primer was able to exclude *CYP2D8P* also, since the primer mismatched at penultimate and antepenultimate bases due to variation in the pseudogene sequence. The primer has more than one mismatch in the pseudogenes, but only one in the active gene allowing it to extend in the active gene only.

4.3.1.2.1 Materials and methods

The reaction was carried out in a thermocycler and the product size was determined by gel electrophoresis. MCA on the Lightcycler was run to determine the Tm of the thermocycler ASPCR product.

i) ASPCR protocol: 25µl of PCR mix was prepared with 2µl of 100ng/µl purified genomic DNA sample, 1µl of each 10µM primer CYP3 paired with either WR42G or MR4+9 in two different reactions for each genomic sample (Integrated DNA Technologies, San Diego, CA) (Table 4.2), 5 µl of each buffer A and buffer B

(Invitrogen, CA, USA), 0.75µl of 50mM $MgCl_2$, 0.5µl of 10mM dNTPs, 2µl Elongase Enzyme Mix (Invitrogen, CA, USA) and 6.75 µl of distilled water. The PCR cycling parameters were the following: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s, annealing and extension at 70°C for 6min and iii) final extension of 70°C for 10min.

ii) Gel electrophoresis was done to confirm the size of the long range PCR product. The PCR products were run on 1% Agarose for 1 hour at 100V against a 10kb ladder.

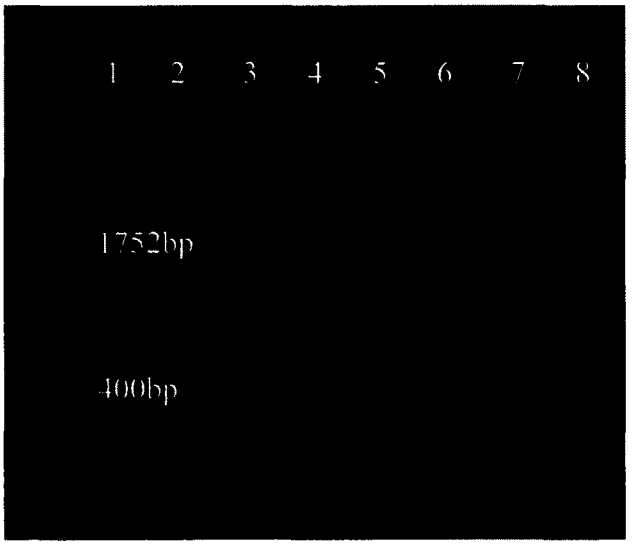
iii) 2µl of 10xLC Green Plus (Idaho Technology Inc., Salt Lake City, Utah) was added to the 18µl of each PCR product and melt curve analysis was performed on Lightcycler 2.0 (Roche Applied Science, USA) from 50°C to 99°C.

4.3.1.2.2 Results

The ASPCR reaction results were promising with the reduced PCR product of 1.7kb base pairs (Figure 4.4). However, the MCA (Figure 4.5) did not yield a sharp melting peak for the product. The potential reason behind it and the solution is discussed in the next section. From the pool of samples, each one of the homozygous wild type and heterozygous did not have sufficient DNA to be

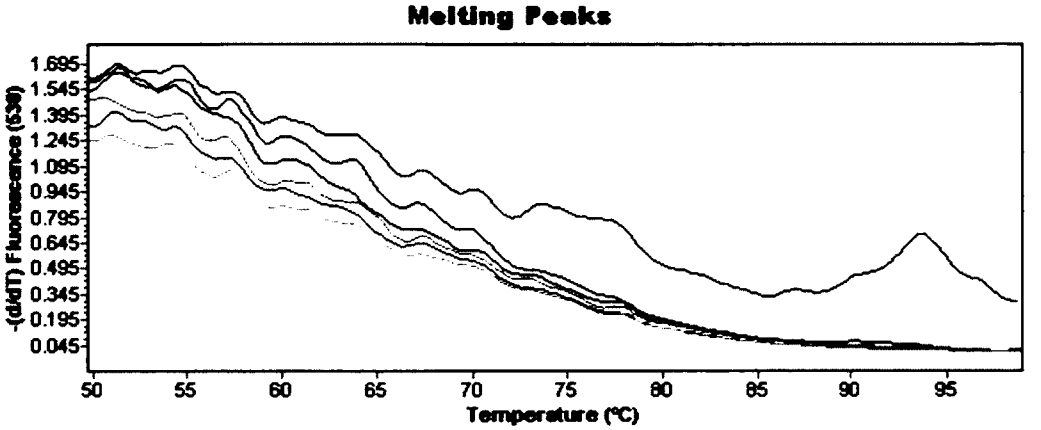
tested, hence they did not yield ASPCR result. The reaction successfully detected 7/8 wild type (wt/wt) and 4/4 mutant (mt/mt) samples and 13/14 heterozygous (wt/mt) samples. Failed samples had insufficient DNA for the test, so of evaluable samples, all were correctly identified. The ASPCR product size is still large enough to cause secondary product formation during the reaction in few of the samples.

Figure 4.4 Gel electrophoresis of ASPCR product from strategy I in genomic DNA



Strategy II has reduced the allele specific PCR product size to 1752bp. Lanes 1-4 are primer-DNA matched product of size 1752bp (lane 1=mutant type primer with mt/mt sample, lane 2=mutant primer with heterozygous sample, lane 3= wild type primer with wt/wt sample, lane 4= wild type primer with heterozygous sample). Even strategy II has shown a secondary unintended product amplification of size ~400bp. Lanes 5 and 6 are the primer-DNA mismatches with wild type primer and mutant primer respectively. Lanes 7 and 8 are the no-template control with the wild type and mutant primer respectively.

Figure 4.5 The melt curve analysis of long range ASPCR product in genomic DNA



The MCA graph shows the melt curve for the wild type primer with six wild type genomic samples. The melt curve analysis of ASPCR product of size 1752 bp did not produce a sharp melting peak upon MCA. The above result shows the melt curve for wild type primer with wt/wt sample as an example of melting of large PCR product.

4.3.1.2.3 Challenges

Strategy II yields a 1752 base pair allele specific product in the conventional system. However, the large DNA duplex is a stretch of multiple loops of double stranded DNA instead of a single long stretch of duplex DNA. The double stranded DNA of larger size (>1kb) melts by segments across the melting temperature with very broad melt peaks [2, 3]. Therefore, a sharp melt profile cannot be generated for the PCR product, since the bases with rich AT segments will melt at the start of the melt and the GC rich segments towards the end, and the segments having nearly equal proportion of GC melt at the centre of the set range of melting temperatures. Hence, I needed to solve three problems.

1) The pseudogenes have to be excluded during the ASPCR. 2) The ASPCR product has to be <500bp in order to generate a sharp melt profile, which gives the status of the allele (wild or mutant or heterozygous). 3) The need to eliminate secondary product formation during the long range ASPCR. These challenges could only be met with Tetra-Allele Refractory Mutation System (Tetra-ARMS) PCR. The secondary product formation could only be reduced by redesigning the long range PCR primers.

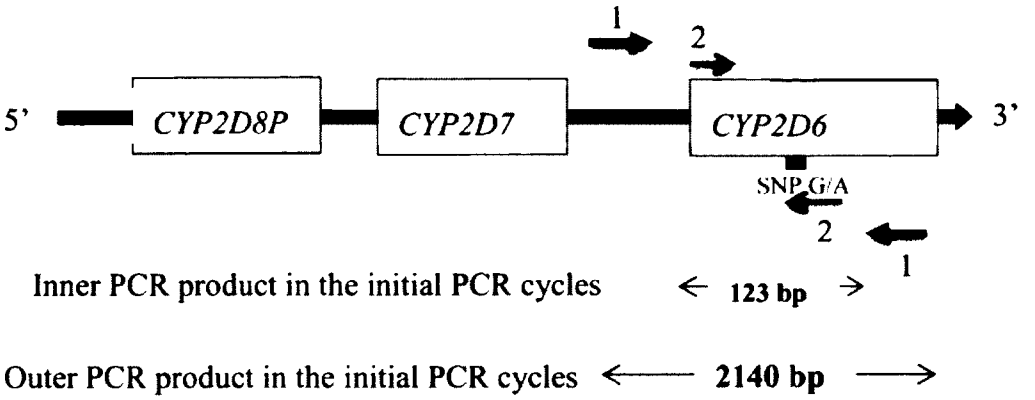
4.3.2 Tetra-ARMS PCR Strategy

The strategy includes the use of a set of two primer pairs within an individual reaction. The first primer pair of the set was used as external primers and the other pair as internal allele specific primers (a "nesting" strategy). The external primer pair was employed for the initial few PCR cycles, and it was designed to bind to the regions specific to the active gene. Therefore, it only amplifies the active gene and excludes the pseudogenes during the PCR. After a few cycles, the amplified active gene copies will be abundant and thus readily available to the inner primer pair. The internal allele specific primer pair (wild or mutant) was employed to perform ASPCR (Figure 4.6). The outer primer pair has an annealing temperature greater than the melting temperature of the inner primer pair (Table 4.3). The primer pairs were designed such that the outer primers have annealing temperatures above the melting temperature of the inner primer pair. During the first few PCR cycles the annealing temperature was set at the optimum for the outer primers, at which inner primers cannot anneal to the DNA. The PCR from the first few cycles targeted only the active gene, and the PCR product was

used as a source of target DNA in the subsequent cycles to perform ASPCR. In the subsequent PCR cycles, the annealing temperature was reduced to favor the binding of inner primers to the amplicon generated from the outer primers (active gene copies produced in the first cycles).

At the end of the PCR, there were two PCR products (large sized product from the outer primers and the allele-specific product from inner primers), which were confirmed by agarose gel electrophoresis. When the PCR products were analyzed by MCA, due to the discontinuous and broad melting nature of the larger product, only the smaller allele specific products show a sharp melting peak, thereby allowing identification of the status of the SNP allele.

Figure 4.6 Tetra-ARMS PCR



Two primer pairs (1 and 2) were used in a single reaction. Pair 1 being the outer primer pair was engaged only in the first few PCR cycles to amplify the active gene *CYP2D6* and exclude the pseudogenes during the PCR. In the subsequent cycles, the inner primer pair 2 performs allele specific PCR. The target for ASPCR is the amplified active gene from the first PCR cycles. Further, MCA was run on the Lightcycler for each PCR product to confirm the melting temperature of the product.

4.3.2.1 Materials and methods

To validate the inner allele specific primers, an ASPCR reaction was carried out using plasmid DNA that contains target sequence from active gene only. Later, the strategy is taken to genomic DNA in a tetra-ARMS format where four primers were used in a single reaction.

4.3.2.1.1 ASPCR in conventional system using plasmid DNA

In total eight reactions were carried out. Each four of the reactions were run with wild type primer (primer-pDNA match for wt/wt, match for wt/mt and mismatch for mt/mt, primer with no-template control) and other three reactions

with mutant type primer (primer-pDNA match for mt/mt, match for wt/mt and mismatch for wt/wt, primer with no-template control). The reaction was carried out in a thermocycler and the product size was determined by gel electrophoresis. MCA on the Lightcycler was run to determine the melting temperature of the thermocycler ASPCR product.

i) PCR protocol for testing inner primers ASPCR in plasmid DNA: For each ASPCR reaction 2μl of 25ng/μl plasmid DNA sample was added to the 23μl of reaction mix containing 2.5μl of 10X PCR buffer, 1μl of 50mM $MgCl_2$, 0.5μl of each 10μM forward primer (Fh19-S) and 10μM of wild reverse (WR2-S) or mutant reverse (MR2-S) (Integrated DNA Technologies, San Diego, CA) (Table 4.3), 0.5μl of 10μM dNTPs, 0.5μl of *Taq* 2011 (5U/μl), 5 μl of 5X UB1 buffer, 0.8 μl of 1% BSA and 12.2 μl of distilled water. The ASPCR cycling parameters were i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 15s, annealing at 55°C for 20s and extension at 72°C for 20s iii) final extension of 72°C for 120s. Post ASPCR products were either immediately used to run gel electrophoresis or stored at -4°C for later use.

ii) Gel electrophoresis of ASPCR product in plasmids was performed to confirm the size of the ASPCR product. The ASPCR products

were run on 2% agarose for 55 minutes at 160V. The products were run along with the 100bp DNA reference ladder.

Table 4.3 Primers for tetra-ARMS reaction

S.No	Primer	Primer Sequence	Product Size
Outer primers employed to exclude the pseudogenes in initial PCR cycles			
1	P100	5' ggcctaccctgggtaagggcctggagcagga	2140bp
2	SEQ64R	5'agaccgttggggcgaaaggg	
Inner primers employed to perform ASPCR in subsequent PCR cycles			
3	Fh19-S	5'cggccttcgccaaccact	123bp
4	WR2-S	5'gcgaaaggggcgtac	
5	MR2-S	5'gcgaaaggggcgtat	

P100 pairs with SEQ64R and excludes the pseudogenes by binding to the active gene specific sequence upstream to the SNP region. The product thus obtained with this primer pair during the initial few PCR cycles is used as the template source by the Fh19 primer pair. Fh19-S pairs with either WR2-S (wild type) or MR2-S (mutant type) to make an allele specific primer pair. The annealing temperature of this primer pair is less than the outer primer pair to facilitate a nested PCR condition. The red base in the reverse primer denotes a deliberate mismatch to increase the stringency of the reaction, and the blue base sits at the SNP base.

4.3.2.1.2 Results with plasmid DNA

The plasmid DNA was used as a tool to test the allele specificity of the inner primer group. There was no necessity to exclude the pseudogenes as plasmid DNA contained only the active gene target sequence. The primers detected *4 alleles successfully in the plasmid DNA as shown in Figure 4.7.

Figure 4.7 Gel electrophoresis of tetra-ARMS inner primer pair using plasmid DNA

The inner primer pair of tetra-ARMS successfully performed ASPCR reaction in the plasmid DNA producing a 123bp of product. Lanes 1 and 2 are no-template controls, lanes 3 and 4 are mismatches of wild type and mutant primer, respectively. Lanes 5 and 6 are mutant primer with mutant DNA and heterozygous DNA, respectively. Lanes 7 and 8 are wild type primer with wild type DNA and heterozygous DNA respectively. This result proved the allele specific nature of the inner primer pair of the tetra-ARMS group of primers.

4.3.2.1.3 Tetra-ARMS in genomic DNA in conventional system

Both inner and outer primers were added to the reaction mix. The outer primers remained the same in all the reactions. However, the inner primer pair was different for wild type or mutant type. Two individual reactions were performed for each genomic sample with the inner primer pair being wild or mutant type.

The first 10/35 PCR cycles were performed at a higher annealing temperature to facilitate the annealing and extension of only outer primers for the purpose of active gene amplification, whereas in the subsequent PCR i.e. 25/35 cycles the annealing temperature was lowered to utilize the inner primers to perform an allele specific PCR reaction.

i) Tetra-ARMS PCR protocol: 25µl of the reaction mix contained 2µl of 100ng/µl genomic DNA sample, 2.5µl of 10X PCR buffer, 1µl of 50mM $MgCl_2$, 0.25µl of each 10µM outer primers (P100 and SEQ64R), 0.5µl of each 10µM forward primer (Fh19-S) and 10µM of wild reverse (WR2-S) or mutant reverse (MR2-S) (Integrated DNA Technologies, San Diego, CA) (Table 4.3), 0.5µl of 10µM dNTPs, 0.5µl of *Taq* 2011 (5U/µl), 5 µl of 5X UB1 buffer, 0.8 µl of 1% BSA and 4 µl of distilled water.

For the first 10/35 PCR cycles, the parameters were: i) initial denaturation of 94°C for 120s ii) 35 cycles of denaturation of 94°C for 10s, annealing and extension at 70°C for 6min and iii) final extension of 70°C for 10min. For the rest of 25/35 PCR cycles, the parameters were i) initial denaturation of 94°C for 120s ii) 25 cycles of denaturation of 94°C for 15s, annealing at 55°C for 20s and extension at 72°C for 20s iii) final extension of 72°C for 120s. Post PCR products were either immediately used to run gel electrophoresis or stored at -4°C for later use.

ii) Gel electrophoresis was done to confirm the size of the long range PCR product. The PCR products were run on 1% Agarose for 1 hour at 100V against a 100bp ladder.

iii) 2µl of 10xLC Green Plus (Idaho Technology Inc., Salt Lake City, Utah) was added to the 18µl of each PCR product and melt curve analysis was performed on Lightcycler 2.0 (Roche Applied Science, USA) from 50°C to 99°C.

4.3.2.1.4 Results with genomic DNA

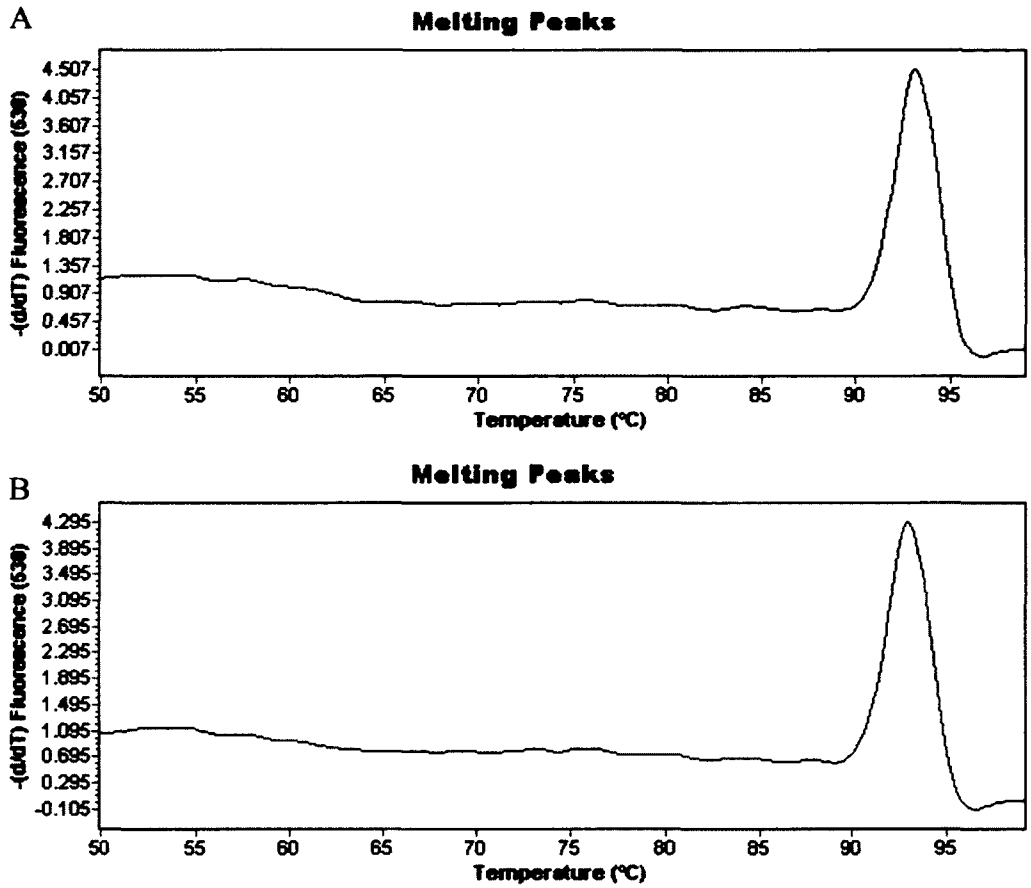
The tetra ARMS PCR strategy was successfully implemented in genomic samples. The tetra-ARMS strategy used outer and inner primer sets. The outer primer set excluded the pseudogene and amplified a product of 2140bp. The product was used as a template by the inner primers to perform ASPCR and

produced a shorter product of 123bp (Figure 4.8). Therefore, the tetra-ARMS test excluded the pseudogene and performed the allele *4 detection test (wild/mutant/heterozygous) with a reduced product size. The smaller product size of 123bp produced a sharp melt peak upon melt curve analysis on the Lightcycler at ~93°C (Figure 4.9 to Figure 4.11). The reaction was successful for 7/8 wild type (wt/wt), 4/4 mutant (mt/mt) samples and 13/14 heterozygous samples (wt/mt). 2/2 deletion samples did not produce any PCR product. 1/8 homozygous wild type and 1/14 heterozygous samples did not have sufficient DNA to be tested; hence they did not show PCR amplification.

Figure 4.8 Gel electrophoresis of tetra-ARMS PCR product using genomic samples

Tetra-ARMS test was successfully implemented in a conventional system using genomic samples. Lanes 1 and 2 show the match of the wild type primer with the wt/wt and heterozygous samples respectively. Lanes 3 and 4 show the match of the mutant primer with mt/mt and heterozygous samples respectively. Lanes 5 and 6 do not show 123 bp PCR products due to the mismatch of wild type primer with mt/mt sample and mutant primer with wt/wt sample respectively. Lanes 7 and 8 are no-template control for the wild type and mutant primers respectively.

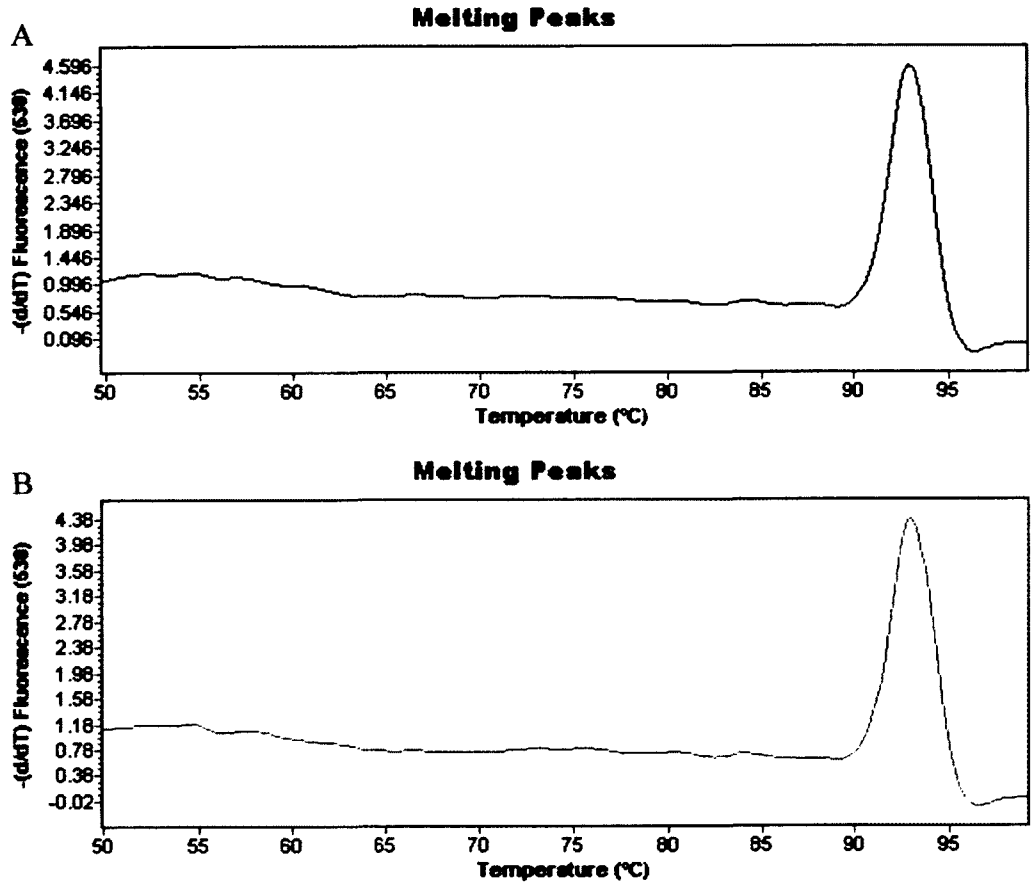
Figure 4.9 MCA of tetra-ARMS product with wild type primer (matched DNA)



A= wild type primer with wild type homozygous sample (wt/wt)

B=wild type primer with heterozygous sample (wt/mt)

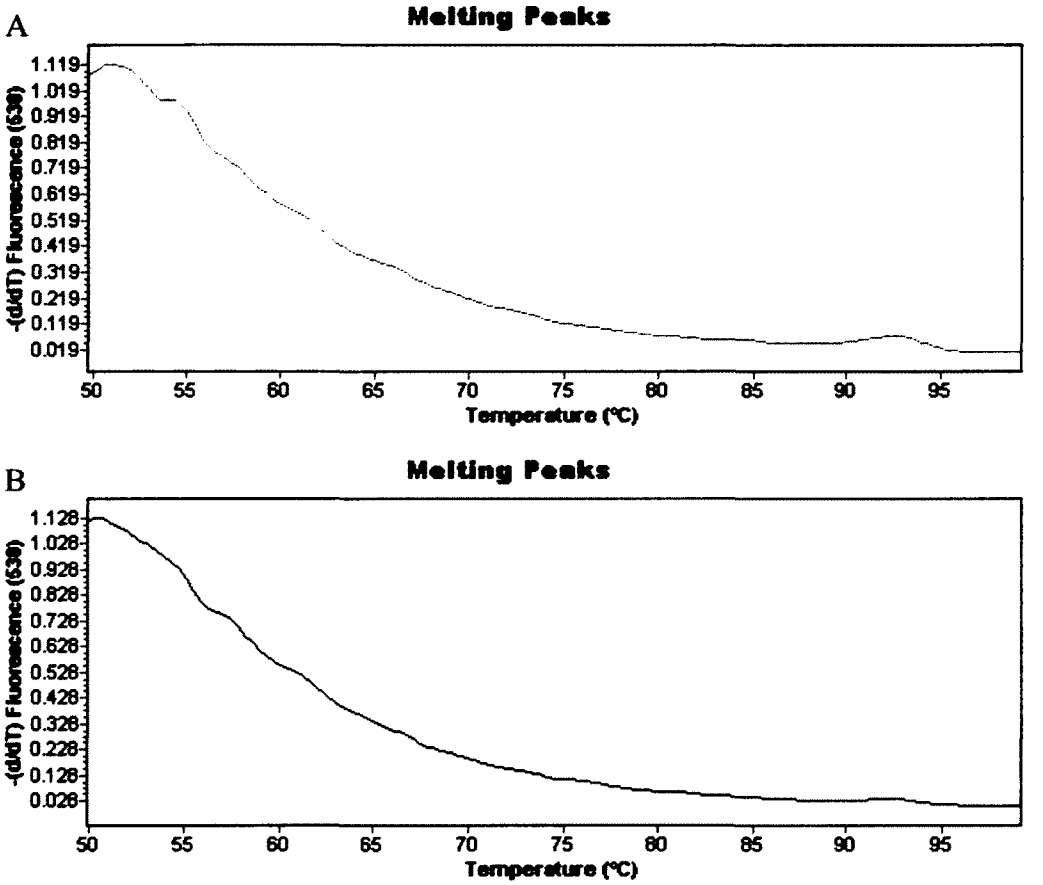
Figure 4.10 MCA of tetra-ARMS product with mutant primer (matched DNA)



A= mutant primer with wild type mutant sample (mt/mt)

B= mutant primer with heterozygous sample (wt/mt)

Figure 4.11 MCA of tetra-ARMS product with wild and mutant primer (mismatched DNA)



A= wild type primer with mutant homozygous sample (mt/mt)

B= mutant primer with wild homozygous sample (wt/wt)

4.3.2.2 Conclusion

CYP2D6 is highly homologous (~97%) to the pseudogenes CYP2D7 and CYP2D8P. Therefore, CYP2D6 is currently genotyped for its SNPs by amplifying a large PCR product where primers bind at the intergenic sequence specific to CYP2D6. The result of such PCR is a long range PCR yielding a product size of 4.7kb. Such a large PCR product is not suitable for gel post application since the amplification in the gel matrix is limited by the porosity of the gel (concentration of the gel).

ASPCR strategy-I excluded the pseudogenes and amplified only the active gene which was confirmed with the PCR product size of 2136 bp along with some secondary product. To reduce the size of the PCR, ASPCR strategy II was proposed that reduced the PCR product size to 1752 bp. However, the PCR product was too large to produce a sharp melting peak for the allele specific PCR product. This led to the development of a tetra-ARMS test with the genomic DNA that used two outer primers to amplify only the active gene and could be further used by the inner allele specific primers as a template.

In tetra-ARMS strategy, I used one of the primers specific to the CYP2D6 (binds at the intergenic region at the 5' end of the sequence), and the reverse primer (binds to the sequence flanking to the SNP at 3' end) to amplify the PCR product of size 2.1 kb and avoided pseudogene amplification. The amplified 2.1 kb product was a CYP2D6 sequence carrying the allele *4. Further, an inner primer pair was used as an allele specific primer pair to perform ASPCR with the

2.1 kb sequence to yield a product of size 123bp. All four primers (2 outer and 2 inner) were included in the same reaction mix. The first few PCR cycles employed the outer primer to exclude the pseudogenes followed by the next set of PCR cycles with inner primers for ASPCR reaction. While the outer primers were engaged in PCR reaction in the first few cycles, inner primers remain unbound to the target DNA and did not involve in PCR reaction. This helped in avoiding potential binding of the inner primers to the pseudogene sequence while the active gene sequence was amplified. This was accomplished by designing shorter inner primer pair so that they reach melting temperature when the outer primers anneal to the target sequence. The inner allele specific primers made a product of 123 bp which produced a sharp melting peak during melt curve analysis. The tetra-ARMS PCR works efficiently as an allele detection strategy for the target gene where contamination with pseudogenes is challenging in developing a diagnostic test.

References

[1] C. Sachse, J. Brockmöller, S. Bauer and I. Roots, "Cytochrome P450 2D6 variants in a Caucasian population: Allele frequencies and phenotypic consequences," *Am. J. Hum. Genet.*, vol. 60, pp. 284-295, 1997.

[2] H. A. S. D. Poland, "Theory of helix-coil transitions in biopolymers," 1970.

[3] R. M. Wartell and A. S. Benight, "Thermal denaturation of DNA molecules: a comparison of theory with experiment," *Physics Reports,* vol. 126, pp. 67-107, 1985.

Chapter 5. Discussion

CYP2D6 is an important predictor of an individual's response to tamoxifen drug therapy. Different *CYP2D6*4* allele SNP combinations show varied degrees of response to the drug. For example, a heterozygous allele carrier would have better efficiency of drug metabolism than homozygous mutant allele carrier. "On the spot" identification of the patient's genotype would help physicians to make rapid changes in drug dose and drug therapy depending upon the efficiency of a patient's drug metabolism. Currently it is difficult to perform SNP detection in a clinical setting primarily due to high cost, need for skilled staff and use of batch processing.

In our lab, we have developed the gel-post system (integrated PCR/MCA). This approach does not require a DNA sample preparation step. Raw clinical sample in the form of buccal swabs or blood can be applied on the posts, where DNA diffuses into the posts through gel pores and PCR product detection occurs in situ. Therefore, the gel post method would be a low cost, seamless approach compared to conventional genotyping systems. Moreover, prepared gel posts might be desiccated and preserved for long periods and might be easily carried to different locations for genotyping clinically important SNPs using instrumentation designed for automation. The developed gel post system will then only require addition of raw sample onto the posts.

On gel posts, amplification is carried out by allele specific PCR (ASPCR). The amplified target sequence can be analyzed for the presence of the *CYP2D6*4* SNP by heating and cooling the PCR products on the Viriloc system, within the same posts without any intervening manual interventions. This approach has so far proven to be robust for performing melt curve analysis. Therefore, developing PCR/MCA methods on gel posts would provide a prototype for clinically important SNP detection in a seamless fashion. The gel post system would simplify the genotyping process for detecting important SNPs and will serve as a sensitive and cost effective diagnostic platform.

The conventional setting involves PCR on a thermocycler, melt curve analysis on the Lightcycler, and electrophoresis in agarose gel to show the right size PCR product. The results from the gel post setting include gel post PCR and MCA on Viriloc, confirmed by the melting temperature of the PCR product. Once the allele specific PCR is optimized on conventional instruments, it can be tested in the gel post system on Viriloc with all requisite controls.

I started my allele specific primer work with plasmid DNA that was cloned to contain the SNP and the flanking gene sequence. I then successfully designed ASPCR primers that amplified CYP2D6 from plasmid templates in conventional and in the gel post systems. However, when the same primers were used with genomic DNA as template, they failed to produce a PCR product, since they were too short to be specific to the genomic DNA sequence. The primer length was increased by 4 bases at the 5' end, and a PCR product was obtained. Unfortunately, the new primers with the added bases produced the correct sized

PCR product but failed to show allele specificity, likely due to their binding to the highly homologous sequences of pseudogenes [1].

I then made several different primer designs that were assumed to be specific to the *CYP2D6* gene, only considering the variant bases in the pseudogenes, but which in theory excluded pseudogenes from the amplification process through the design of the common forward primer to the allele specific reverse primers. The variation sites in the active gene are the bases at which the active gene *CYP2D6* differs from the two pseudogenes *CYP2D7* and *CYP2D8*, and the primer binds at the variation site with its 3'end. I tested them with the ASPCR reverse primer in conventional PCR and in the gel post system. However, primers failed to show allele specificity in the conventional system as well as in the gel post system for mismatched samples. The reason underlying this failure appears to be that the primers were not able to differentiate between active and pseudogenes. It might be due to the inconsistency of the variant bases in the gene sequence, which are at the 3' end of the primers. Therefore, the variation site might not be the correct location for designing primers to exclude the pseudogenes during ASPCR. Multiple primer sets were tested, most of which failed to exclude pseudogenes. The details of all the failed primer designs are provided in the Appendix.

Before designing any primers for ASPCR, I proposed to I) screen a set of 28 samples for single gene, deletion and duplication events in order to avoid any false negative results during test development, and II) screen the population for the SNP status by excluding the pseudogenes. For the first work, a published set

of primers were taken [2-4]. To identify the SNP, I used a 2 step PCR form. The first step was a long range PCR to exclude the pseudogene, with this long range PCR product used as template for the 2nd reaction generating a shorter product carrying the SNP of interest. Based on sequencing of the amplicons, out of 28 genomic samples only 2 samples were found to have deletions; the rest were all single gene carriers. The SNP screening of the population allowed me to use these samples as controls with a known genotype for developing the ASPCR test. Although complex and labor intensive, development of this set of samples with known genotypes was a critical advance that ultimately allowed me to accurately verify PCR results designed to exclude pseudogene amplification. Using this strategy, the tetra-ARMS approach was found to successfully identify the SNP of interest and effectively exclude pseudogenes.

To develop the working ASPCR test, I proposed two ASPCR strategies. The two strategies have similar reverse allele specific primers (wild or mutant type). However, the two strategies differ in the common forward primer. The first strategy makes use of a primer that binds to the intergenic region between the pseudogene CYP2D7 and the active gene. Whereas for the second strategy, I considered the 'mutation site' that turned an active gene into a pseudogene as a binding site for the common forward primer. In both the strategies, the forward primer ensures that the sequence is *CYP2D6*, while the reverse primer extends only in the presence of wild/ mutant match DNA.

Strategy-I produced a PCR product of 2136bp. This led to the proposal of the second strategy that reduced the PCR product size to 1752bp. I was able to

exclude the pseudogene and carry out an allele-specific reaction, but the limitation for analysis was the size of the PCR product along with the contamination of the reaction with secondary product amplification.

The challenges were to develop a test using allele specific primers that can give a sharp melting peak and still exclude the pseudogenes. The sharp melting peak is possible only with the short PCR products. This led to the idea of tetra-ARMS PCR strategy that helped in shortening the final ASPCR product so MCA became possible and the reaction can then detect the presence or absence of an ASPCR product.

The tetra-ARMS PCR test employed 2 primer pairs, one for the outer region of the gene employed during the first 10 cycles to generate a product of size 2140bp, whereas the subsequent 25 cycles were carried out with an inner primer pair that produced a 123bp amplicon in a nested PCR format. The first primer pairs were designed with annealing temperatures above the melting temperatures of the inner primer pair. Therefore, the inner primer pair could not bind to the target for the first few cycles during which the pseudogene is excluded in the PCR product. In the subsequent PCR cycles, the annealing temperature is reduced allowing the inner primers to amplify only the active gene as the target for allele specific PCR. Furthermore, the 123bp product was short enough to provide us with a defined melting temperature profile of the ASPCR product.

The novel approach in my work is the reduction of the PCR product size to 2.1 kb specific to CYP2D6 that might be applicable to the gel post system. My

strategy thus avoids the amplification of highly homologous (97%) pseudogenes of CYP2D6 while accommodating the requirement of the gel post system for a small product. Therefore, by employing the tetra-ARMS strategy pseudogene amplification can be avoided and the ASPCR can be performed to detect the allele *4 with a much smaller detectable product size (123bp), which is important when considering the test implementation on gel post system.

In the gel post system, the tetra-ARMS approach will make use of at least six sets of replicate gel-posts run simultaneously in a single array setting, as illustrated in (Figure 5.1).

The developed test in Figure 5.1 could be used for detecting **4* allele, Post array # 1 2 lacks template (no-template control NTC) will have wild type or mutant type primer, and they will not generate any melt curves. Post arrays # 3 and 4 will give a melt curve after PCR since each of them have wild or mutant specific primer with the wild or mutant DNA template respectively, as positive controls. Post arrays # 5 and 6 have all reaction reagents except DNA, and they are used to test the clinical raw samples. Externally, the raw sample containing DNA is applied only to post arrays #5 and 6.

The post arrays #5 and 6 are the test-posts, where the DNA sample is applied externally. They lack template until the DNA sample is applied to them. If both of them generate product as assed by the MCA, they will be identified as heterozygous for **4* allele. If one of test post arrays but not the other generates product verified by MCA, then it is homozygous for the wild type or mutant allele

depending on which primer set amplifies a product. Table 5.1 shows the output of the **4* SNP test outcomes using the **4* tetra-ARMS test on the gel post system. It shows that the test on gel posts could reveal the patient status as a poor metabolizer or extensive to ultra-rapid metabolizer. Therefore, it can be used in the clinical setting to predict the tamoxifen metabolism status of the patient, and the approach would serve as a prototype to detect panels of SNPs important for pharmacogenetic studies.

Whether CYP2D6 genotyping is related to tamoxifen metabolism remains controversial. Here, I analyze the outcome of six published trials. All studies were retrospective follow-up studies carried out in Caucasian population in adjuvant treatment setting. Three of the studies [5-7] were consistent with the hypothesis that poor metabolises of tamoxifen have higher recurrence rate compared to extensive metabolisers. However, other studies[8-10] failed to show the association between genotype and drug response.

Drawing conclusions based on these studies is difficult, since the studies with contradictory results had limitations in their study design. For instance, they did not consider the co-medication of CYP2D6 inhibitors in their study designs[8, 9]. Moderate to severe hot flashes occur in the patients with wild type allele (extensive metaboliser); as a result patients are administered CYP2D6 inhibitors (selective serotonin inhibitors). Administration of CYP2D6 inhibitors thus decreases the drug metabolism in patients with extensive metabolism genotype. Not considering co-medication could incorrectly lead to the conclusion that there is no co-relation between the genotype and response to the tamoxifen therapy.

Furthermore, issues around patient compliance were not addressed. Patients who respond well to tamoxifen (good metabolizers) often have more severe side effects which is known to result in reduced compliance by some of the patients who chose to avoid side effects by failing to take their medication.

Some of the studies included patients carrying heterozygous allele along with the homozygous wild type allele [5] or homozygous mutant allele [6-10]. Moreover, there is not enough data to confirm if the average endoxifen level in the heterozygous patients is enough to achieve the clinical response to the drug. The level of expression in the heterozygous patients varies depending upon the extent of function allele expression, hence if the heterozygous carriers can be either extensive metabolisers or intermediate metabolisers. Therefore, validating the hypothesis and drawing conclusion from the published studies might be very difficult.

Studies were carried out in Asian subjects to relate the genotype with the response to tamoxifen [11-14]. The homozygous mutant allele (CYP*10) in all the studies supported the hypothesis that poor metaboliser phenotype had less benefit from tamoxifen therapy. To conclude, from the work published so far, the relation of CYP2D6 alleles specially *4 with the response to tamoxifen is clear. However, further clinical trials are needed to strengthen this relation. Therefore, in general genotyping CYP2D6 holds great promise in personalized medicine for tamoxifen response in breast cancer patients when it is genotyped along with the other alleles involved in tamoxifen metabolism.

The four SNPs *3, *4, *5 and *6 are responsible for 97% of all the alleles causing PM in the Caucasian population. However, the *4 allele is the most frequent allele in the Caucasian population, and it is the major contributors to the PM phenotype. Genotyping allele *4 provides us with a good prediction of drug metabolizing status, yet is insufficient to fully understand the status of patient every time without knowing the status of *3 and *5 (deletion) and *6 alleles. These alleles should be tested along with the *4 allele, and the strategy of tetra-ARMS test for detecting CYP*4 allele can be used to develop the test for other alleles.

Allele *4 test itself should not be considered as the sole contributor to an evaluation of the drug response in patients. Other factors should be evaluated that include patient's age, liver function status, hepatic blood flow, renal drug elimination, as they can alter drug metabolism and elimination. Patient's ethnic group has to be considered as *4 allele has highest penetrance only in Caucasian population. Occurrence of hot flashes, estrogen hormone receptor status and stage of the breast cancer are also important as they are clinically related to the response of a patient to tamoxifen therapy and its side effects [15].

The tetra-ARMS PCR has been successfully developed and can now be implemented on the gel post system. In future work, the sensitivity of the test will be analyzed by a serial dilution method. Then we can test the breast cancer patient samples acquired from Dr. Mackey's lab at Cross Cancer Institute, Edmonton, and compare the gel post results with the conventional PCR, sequence them on the ABI3130xl, and compare test results with the actual drug response for each

patient. The positive and negative controls would be integrated on the gel post system when testing the samples from breast cancer patients.

To make the developed tetra-ARMS test clinically useful, the DNA sample in the form of raw blood/ plasma, or potentially use of buccal samples, has to be supplied externally. In this case the gel posts will contain the PCR mix including the primers. The only human intervention will be the addition of the DNA sample on the array of gel posts. Once tested with raw clinical raw samples, the system can provide multiplexed analysis for different SNPs or disease targets, and can be rapidly customized for a wide breadth of diagnostic applications. The test will be a useful tool to guide development of diagnostic tests where challenges might be posed by pseudogenes.

Roche developed AmpliChip CYP450 test using microarray technology from Affymetrix (GeneChip) to determine 27 alleles of the CYP2D6 gene, and the test is approved by FDA. Another company Genelex uses xTAG Luminex analyzer which performs multiplex allele specific extension followed by hybridisation detection strategy that detects 17 alleles of CYP2D6. Their strategy of testing multiple alleles suggests that increasing the number of detected alleles of CYP2D6 involved in drug metabolism on a single gel post chip is important. Targeting multiple alleles on a single chip will help to I) increase the clinical validity of the test ii) speed up the regulatory approvals iii) improve marketability of the product and to increase customer satisfaction. The gel post system is likely to be a better system as the only PoC diagnostic system for CYP2D6 genotyping. Single mutation detection (*4) Tetra ARMS PCR might serve for the purpose.

However, multiple allele detection would need redesigning the assay with more innovative approach, as with the tetra-ARMS approach uses strict PCR conditions which might not be applicable to simultaneously detect other alleles.

Incomplete penetrance of the target SNPs and genetic heterogeneity for the new SNPs on the panel will be a limitation for clinical validity of the test. Hence, in future it is important to select the SNPs that have confirmed positive predictive value.

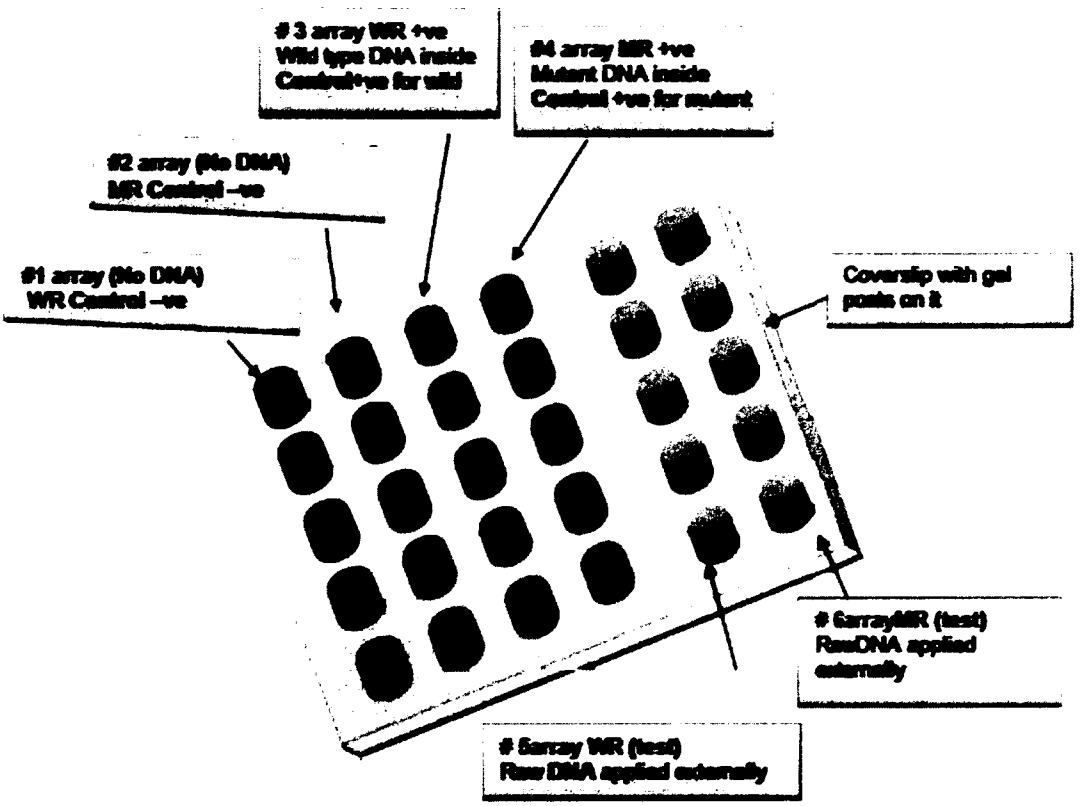
Some of the very important criteria to develop the final gel post system might include I) extracting enough amount of DNA to be tested, II) delivering enough amount of DNA to the gel posts without cross-contamination across the posts, and III) avoiding cross contamination between the positive controls (that are pre-loaded with control DNA). The heating element's (peltier) temperature and the medium properties (wax or oil in which the gel posts are submerged) might alter overtime and change the PCR thermal cycling parameters. Proper thermal calibration methods for peltier are required to create a controlled environment for PCR thermocycling. Getting proper gel posts (~100%) from the mold and mechanical strength and composition (liquid vs. gel) of the gel post are also important as non-polymerised gel might inhibit the PCR, absorption of DNA and or interfere with the fluorescence. These will all be tested in the future.

Maintenance of the gel post system is another crucial aspect to be considered in parallel to developing the gel post system, since it is important for post-sale maintenance and a prerequisite for market approval. Hence, validation

tests have to be performed for the reusability of the gel post systems. Adequate SOPs and protocols are very important for PoC diagnostics, as non-technical staff might have difficulty with the quality control documentation for reliable testing specially for point of care genetic test that targets multiple alleles and demands proper handling of the instrument.

This project can be taken further to develop ASPCR for the remaining important alleles in the Caucasian population i.e. *3, *5 and *6. Nevertheless, *4 allele is highly frequent in the population. Genotyping of these alleles in the gel post system would serve as a cost effective prototype system for the pharmacogenetic testing of other SNPs. The genotyping results will help in making important clinical decisions, in order to reduce the toxicity of the drug and increase its therapeutic effect or to make a choice of alternative drugs to treat cancer.

Figure 5.1 Genotyping *4 on gel posts



Each array has a specific primer set. At least six distinct post arrays are included in a single reaction setting. Green color post arrays for test sample. Red color post arrays serve as +ve control and have DNA in them. The black post array is a negative control, which has either wild or mutant primer but will not have DNA. Externally raw DNA sample is applied to the arrays #5 and

Table 5.1 Different genotype-phenotype outcomes of the **4* allele specific test in the gel post system

Post # 5 WR test post	Post # 6 MR (**4*) test post	Wild allele	**4* allele	Expected genotype	Patient's status
PCR product	PCR product	Y	Y	**4*/A, **4*XN/A, **4*/AXN	EM (less chance for IM)
PCR Product	No product	Y	N	*A/*A	UM (less chance for EM)
No product	PCR product	N	Y	**4*/**4*	PM

PM denotes poor metabolizer of the tamoxifen drug. EM and UM are the extensive and ultra-rapid metabolizer of the tamoxifen drug. WR= wild type reverse primer, MR= mutant reverse primer, Y=yes present, N=not present, A=normal active gene

References

[1] S. Kimura, M. Umeno, R. C. Skoda, U. A. Meyer and F. J. Gonzalez, "The human debrisoquine 4-hydroxylase (CYP2D) locus: Sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene," *Am. J. Hum. Genet.*, vol. 45, pp. 889-904, 1989.

[2] C. Sachse, J. Brockmöller, S. Bauer and I. Roots, "Cytochrome P450 2D6 variants in a Caucasian population: Allele frequencies and phenotypic consequences," *Am. J. Hum. Genet.*, vol. 60, pp. 284-295, 1997.

[3] V. M. Steen, O. A. Andreassen, A. K. Daly, T. Tefre, A. Brresen, J. R. Idle and A. Gulbrandsen, "Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology," *Pharmacogenetics,* vol. 5, pp. 215-223, 1995.

[4] E. Lundqvist, I. Johansson and M. Ingelman-Sundberg, "Genetic mechanisms for duplication and multi-duplication of the human CYP2D6 gene and methods for detection of duplicated CYP2D6 genes," *Gene,* vol. 226, pp. 327-338, 1999.

[5] M. P. Goetz, J. M. Rae, V. J. Suman, S. L. Safgren, M. M. Ames, D. W. Visscher, C. Reynolds, F. J. Couch, W. L. Lingle, D. A. Flockhart, Z. Desta, E. A. Perez and J. N. Ingle, "Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes," *J. Clin. Oncol.*, vol. 23, pp. 9312-9318, 2005.

[6] W. Schroth, L. Antoniadou, P. Fritz, M. Schwab, T. Muerdter, U. M. Zanger, W. Simon, M. Eichelbaum and H. Brauch, "Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes," *J. Clin. Oncol.*, vol. 25, pp. 5187-5193, 2007.

[7] S. Gonzalez-Santiago, R. Zárate, J. Haba-Rodríguez, A. Gómez, E. Bandres, S. Moreno, P. Borrega, J. García-Foncillas and E. Aranda, "CYP2D6* 4 polymorphism as blood predictive biomarker of breast cancer relapse in patients receiving adjuvant tamoxifen," *Journal of Clinical Oncology,* vol. 25, pp. 590, 2007.

[8] S. A. Nowell, J. Ahn, J. M. Rae, J. O. Scheys, A. Trovato, C. Sweeney, S. L. MacLeod, F. F. Kadlubar and C. B. Ambrosone, "Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients," *Breast Cancer Res. Treat.*, vol. 91, pp. 249-258, 2005.

[9] P. Wegman, L. Vainikka, O. Stål, B. Nordenskjöld, L. Skoog, L. E. Rutqvist and S. Wingren, "Genotype of metabolic enzymes and the benefit of tamoxifen in

postmenopausal breast cancer patients." *Breast Cancer Res.*, vol. 7, pp. R284-290, 2005.

[10] P. Wegman, S. Elingarami, J. Carstensen, O. Stål, B. Nordenskjöld and S. Wingren, "Genetic variants of CYP3A5, CYP2D6, SULT1A1, UGT2B15 and tamoxifen response in postmenopausal patients with breast cancer," *Breast Cancer Res.*, vol. 9, 2007.

[11] Y. Xu, Y. Sun, L. Yao, L. Shi, Y. Wu, T. Ouyang, J. Li, T. Wang, Z. Fan, T. Fan, B. Lin, L. He, P. Li and Y. Xie, "Association between CYP2D6 *10 genotype and survival of breast cancer patients receiving tamoxifen treatment," *Ann. Oncol.*, vol. 19, pp. 1423-1429, 2008.

[12] K. Kiyotani, T. Mushiroda, M. Sasa, Y. Bando, I. Sumitomo, N. Hosono, M. Kubo, Y. Nakamura and H. Zembutsu, "Impact of CYP2D6*10 on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy," *Cancer Sci.*, vol. 99, pp. 995-999, 2008.

[13] H. Lim, H. J. Lee, K. S. Lee, E. S. Lee, I. Jang and J. Ro, "Clinical implications of CYP2D6 genotypes predictive of tamoxifen pharmacokinetics in metastatic breast cancer," *J. Clin. Oncol.*, vol. 25, pp. 3837-3845, 2007.

[14] B. Bonanni, D. Macis, P. Maisonneuve, H. A. Johansson, G. Gucciardo, P. Oliviero, R. Travaglini, M. G. Muraca, N. Rotmensz, U. Veronesi and A. U. Decensi, "Polymorphism in the CYP2D6 tamoxifen-metabolizing gene influences clinical effect but not hot flashes: Data from the Italian tamoxifen trial [1]," *J. Clin. Oncol.*, vol. 24, pp. 3708-3709, 2006.

[15] J. E. Mortimer, S. W. Flatt, B. A. Parker, E. B. Gold, L. Wasserman, L. Natarajan and J. P. Pierce, "Tamoxifen, hot flashes and recurrence in breast cancer," *Breast Cancer Res. Treat.*, vol. 108, pp. 421-426, 2008.

Chapter 6. Appendix

6.1 Abandoned PCR strategies proposed for *CYP2D6*4* detection

An allele Specific PCR approach was successfully accomplished using plasmid DNA on both the conventional and gel post systems. The same primers did not yield ASPCR on genomic DNA due to the challenges posed by *CYP2D6* pseudogene sequences.

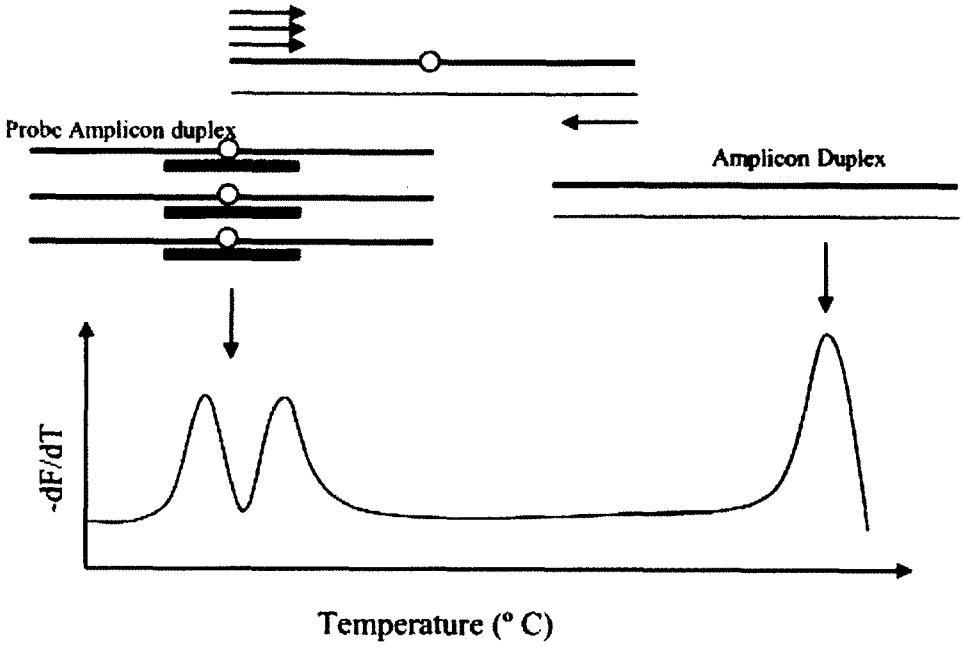
Here I list all the primers that I have designed to accomplish the detection of *CYP2D6*4* SNP rs3892097 status (wild type/ heterozygous/ mutant type) in genomic DNA. I tried several types of PCR assays. They were 1) Asymmetric PCR approach 2) ASPCR approach 3) Tetra primer amplification refractory mutation system (ARMS) PCR 4) Two-way ASPCR approach. All of the following primer designs were abandoned due to their binding to both active and pseudogenes. Later, the active gene-specific ASPCR approach was established. At the end of the Appendix I, LNA based method of genotyping is added which was able to exclude the pseudogene during the PCR, but it was not sufficiently efficient to detect the SNP status.

6.1.1 Asymmetric PCR

The primers for the asymmetric approach were ordered (Table 6.1) before I proposed the idea to implement ASPCR test to test the SNP **4*. I thought to

implement the asymmetric PCR approach where these primers were meant to be used in combination with a probe to detect the presence of the SNP. First, the asyF and asyR amplifies the target sequence carrying the SNP, and then the probe binds to the PCR product during melting. The probe has the SNP base in the sequence, and would completely match the sample PCR product if there is a wild type SNP. Upon melting, in the presence of the wild type SNP, the probe melts at higher temperature rather than when the probe has a mismatch with the mutant allele in the PCR product (Figure 6.1). The *CYP2D6* sequence is a GC rich sequence, and the theoretical difference of Tm (melting temperature) between matched probe-DNA duplex and mismatch was below 2C. This approach was abandoned, since it was difficult to distinguish between the wild and mutant alleles for such a narrow temperature difference on prototype systems (Viriloc) where there were several variables such as temperature gradient across the peltier, imaging the fluorescence etc. These issues within the Viriloc could be solved. The GC rich sequence flanking to the target SNP *CYP2D6*4* rs3892097 is shown in Figure 6.2 along with the primer binding regions. Therefore, I switched my experimental strategy to allele specific PCR (ASPCR). It is to be noted that the AsyF and AsyR were not *CYP2D6* specific, since they could amplify the pseudogene sequences, as seen in BLAST results (Figure 6.3).

Figure 6.1 Asymmetric PCR



One of the primers was added in excess to produce excessive DNA strands complimentary to the probe. Probe is present within the master mix, and produces probe amplicon melt profile differently for wild or mutant SNP. The amplicon duplex is melted at higher temperatures. Here the circle in the solid line is the SNP base.

Figure 6.2 Asymmetric primers and target sequence

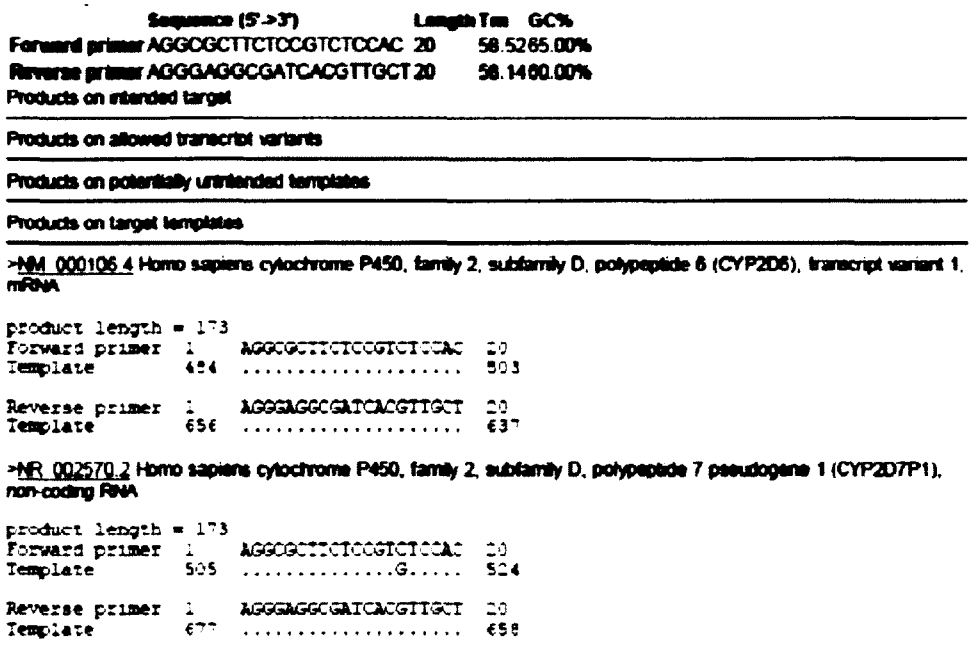
Asy F

aggcgcttctccgtctccaccttgcgcaacttgggcctgggcaagaagtcgctggagcagtgggtgacc
gaggaggccgcctgcctttgtgccgccttcgccaaccactccggtgggtgatgggcagaagggcacaaagcggga
actgggaaggcgggggacggggaaggcgacccccttacccgcatctcccacccccaGgacgcccctttcgcccca
acggtctcttggacaaagccgtgagcaacgtgatcgcctccct

Asy R

AsyF and AsyR were the primers designed to do asymmetric PCR, which were not used later due to the asymmetric PCR strategy limitation to resolve between mutant and wild type alleles and also the possibility of giving a false results by binding to pseudogenes along with the active gene. The bold letter is the SNP position (G=wild type base / A= mutant base)

Figure 6.3 AsyF and AsyR picked up pseudogene upon BLAST



	Sequence (5'->3')	Length	Tm	GC%
Forward primer	AGGCGCTTCTCCGTCTCCAC	20	58.52	65.00%
Reverse primer	AGGGAGGCGATCACGTTGCT	20	58.14	60.00%

Products on intended target

Products on allowed transcript variants

Products on potentially unintended templates

Products on target templates

>NM_000106.4 Homo sapiens cytochrome P450, family 2, subfamily D, polypeptide 6 (CYP2D6), transcript variant 1, mRNA

```
product length = 173
Forward primer  1    AGGCGCTTCTCCGTCTCCAC  20
Template        484  ....................  503

Reverse primer  1    AGGGAGGCGATCACGTTGCT  20
Template        656  ....................  637
```

>NR_002570.2 Homo sapiens cytochrome P450, family 2, subfamily D, polypeptide 7 pseudogene 1 (CYP2D7P1), non-coding RNA

```
product length = 173
Forward primer  1    AGGCGCTTCTCCGTCTCCAC  20
Template        505  ..............G.....  524

Reverse primer  1    AGGGAGGCGATCACGTTGCT  20
Template        677  ....................  658
```

AsyF and AsyR picked up pseudogene sequence upon performing BLAST along with *CYP2D6* gene, hence they were not specific to *CYP2D6* gene.

Table 6.1 Primers for asymmetric PCR

S.No	Primer/Probe	Primer Sequence	Product Size
1	AsyF	5'aggcgcttctccgtctccac	261 bp
2	AsyR	5'agggaggcgatcacgttgct	
3	Probe 1	5'cccacccccaggacgcccctttc	23 bp
4	Probe 2	5'atctcccacccccaggac	18bp
5	Probe 3	5'ctcccacccccaggacgcccctttc	25 bp

AsyF and AsyR were designed for the Asymmetric PCR approach, which was abandoned because of the melting temperature resolution issues. The probes were meant for post-PCR detection of the SNP allele (wild type/ heterozygous/ mutant type). The probes 1, 2 and 3 differ in their length and the position of the SNP.

6.1.2 Allele specific PCR

The allele specific primers which were successful on plasmid DNA did not give actual product when using the genomic DNA. Four bases were added to the primers at the 5' end to increase the PCR productivity (primer 1-3 in (Table 6.2). They already had a deliberate 2nd base mismatch to increase the specificity of the allele specific reaction. In the conventional setting, these primers showed a single PCR product after adding the extra bases. However, they also showed false positive (in the presence of mismatch) with genomic DNA. The possible problem for these primers not being allele specific could be the presence of pseudogenes,

as pointed out by Dr. Lai in a committee meeting (September 2009). Then I analyzed all the sequences (*CYP2D6*, *CYP2D7* and *CYP2D8PP*) as described by Kimura et al [1]. I noticed that there were certain variations of single bases in the gene sequence where *CYP2D6* is unique when compared to the pseudogenes, and can be used as the binding sites for the common forward primers in order to exclude the amplification of pseudogenes along with the active gene during the PCR or ASPCR. Therefore, I designed the *CYP2D6* specific common forward primers (primers 5-6 in Table 6.2) that do not bind to the pseudogenes as per the sequence comparison between active and pseudogene published by Kimura et al, and could be used in combination with (primers 7-10 in Table 6.2) allele specific reverse primers. Primer 4 in Table 6.2 was designed by the NCBI pick primer tool available online. Among the primers, primer Fh19 and Fh22 (primers 5-6 in Table 6.2) gave better PCR yield than the NCBI generated primer Fpair7 primer (#4). Fh19 was assumed to exclude the amplification of pseudogenes better than Fh22, since it carries two mismatches at the 3' end to the *CYP2D7* and *CYP2D8PP* sequences, whereas the Fh22 has only one mismatch. Primers 7-10 did not give a good PCR product yield for the positive run, and later I added more bases, shown in Table 6.2. However, among all these primers, primers 4 and 5 gave better PCR yield in a conventional PCR test.

There is no evidence that these variation sites are consistent in the Caucasian population. Hence, the use of these primers was abandoned.

Table 6.2 List of ASPCR primers designed to work on genomic DNA

S.No	Primer	Primer Sequence	Product Size
4 bases added to the ASPCR primers that worked in plasmid DNA			
1	F4	5'aggcgcttctccgtctccaccttg	221 bp
2	WR4	5'cgttggggcgaaaggggcgtac	
3	MR4	5'cgttggggcgaaaggggcgtat	
Common Forward ASPCR primers			
4	Fpair7	5'gcctgcctttgtgccgcctt	141 bp
5	Fhomo19	5'gccgccttcgccaaccact	129 bp
6	Fhomo22	5'ggtgatgggcagaagggcacaa	104 bp
Reverse ASPCR primers			
7	WR2A	5'ttggggcgaaaggggcgtac	
8	MR2A	5'ttggggcgaaaggggcgtat	
9	WR2A3G	5'ttggggcgaaaggggcggac	

10	MR2A3G	5'ttggggcgaaaggggcggat

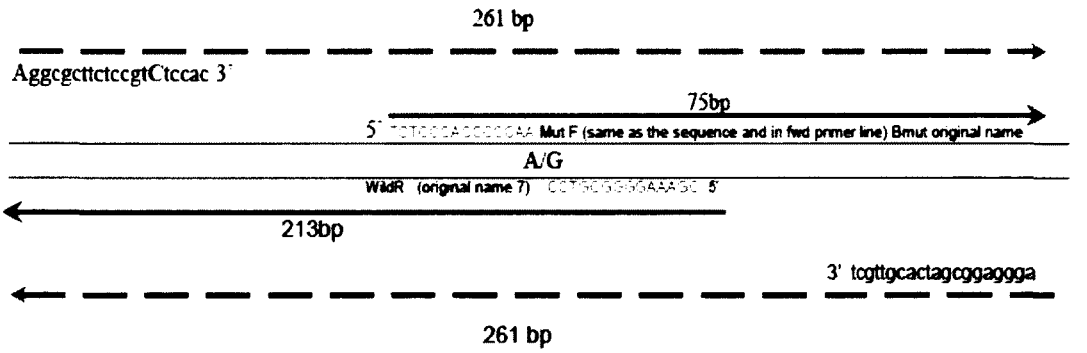
Legend to table 6.2. Primers 1-3 were the working ASPCR primers for plasmid with added 4 bases at 5' end. 4-6 were designed to exclude pseudogene by mismatching with variation sites at 3' end, and hence amplify only the active gene*CYP2D6* (does not amplify pseudogenes*CYP2D7* and *CYP2D8PP*). #7-10 primers were allele specific to ***4*** wild/mutant SNP and work as reverse primers. They were designed to be used along with 4-6 forward *CYP2D6* specific primers. Red coloured base indicates the mismatches introduced deliberately to increase the specificity, whereas the blue indicates the SNP position in the primer. The purple bases in the primers 4-7 were the bases specific to *CYP2D6* active gene only, thus avoiding the amplification of pseudogenes *CYP2D7* and *CYP2D8P*.

6.1.3 Tetra Primer Amplification Refractory Mutation System (tetra-ARMS PCR)

Primers 1-2 in Table 6.3 were meant for tetra-ARMS PCR tests in which a PCR reaction mix has all four primers (two outer and two inner). The two outer primers, in this case AsyF and AsyR from Table 6.1, amplify the region around the SNP in the first PCR reaction. Later in the 2ndPCR reaction, the inner primers (in this case MutF and Wild R) were allele specific and amplified only if they matched the wild type or the mutant SNP. I tried several experiment to optimize this reaction, but abandoned this approach because the AsyF and AsyR (outer primers) were not sequence specific as per Kimura et al., sequencing data and primer-BLAST results. Implementation of tetra-ARMS PCR also needs to set two different PCR reactions one after another with different sets of annealing and

extension temperatures and cycle times (Figure 6.4). Primers 3-6 in Table 6.3 were a modification of WR4 or MR4 (wild and mutant specific primers from Table 6.2) to test the allele specific nature of the primers. Primer 3 in Table 6.3 has two mismatches and does not give a PCR product in a positive run. (Primer 5 in Table 6.3 also has two mismatches around the SNP and does not give a good PCR product yield. Primer 4 in Table 6.3 has better PCR yield than any other wild specific reverse primer in the gel post system even compared to WR4 (Primer 2 in Table 6.2). Primer 7 in Table 6.3 was designed to work as allele specific primer to be used in combination with AsyR (Primer 2 in Table 6.1) since I thought AsyR was *CYP2D6* specific, but later found that it can potentially bind to the pseudogenes as well. Hence, the primer 7 was not used further to test ASPCR.

Figure 6.4 Tetra-ARMS ASPCR



The outer primers (in black) are AsyF and AsyR making *CYP2D6* 261 bp products in the first PCR reaction (red dotted line). The inner primers (in pink) were used in the subsequent 2ndPCR reaction in different thermal cycling conditions to generate allele specific product blue line (75 bp when mutant allele present, 213 bp when wild type allele present). However, AsyF and AsyR were found to be not specific to *CYP2D6* sequence, and possibly could bind to *CYP2D7* and *CYP2D8PP*, so this strategy was abandoned.

Table 6.3 Primers for ASPCR

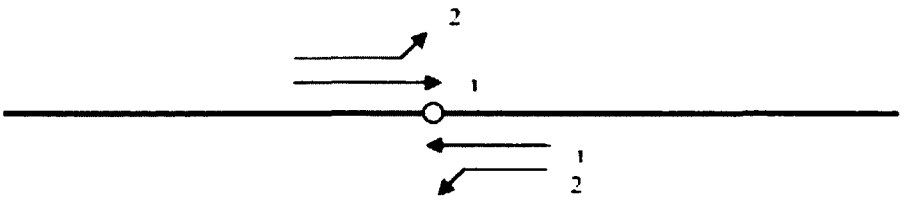
S.No	Primer	Primer Sequence	Product Size
Tetra ARMS PCR primers			
1	Mut F	5'tctcccacccccaa	75 with asyR1
2	Wild R	5'cgaaaggggcgtcc	213 with asyF1
Further modification to ASPCR primer from Error! Not a valid result for table.			
3	WR43G	5'cgttggggcgaaaggggcggac	similar as WR4
4	WR42G	5'cgttggggcgaaaggggcgtgc	similar as WR4
5	WR4-0C	5'cgttggggcgaaaggggcgtacc	add 1 to WR4
6	MR4+2	5'accgttggggcgaaaggggcgtat	add 2 to MR4
ASPCR forward primer to AsyR from Error! Not a valid result for table.			
7	MRF/AsyR	5'acccgcatctcccacccccta3'	83 bp

Primers 1 and 2 were meant to be internal primers to work after the external AsyF and AsyR extend the *CYP2D6* gene in Tetra-ARMS PCR fashion. Primers 3 and 5 were evaluated for the ASPCR in the presence of two deliberate mismatches. Due to the addition of second deliberate mismatch in the primers, they failed to extend during the PCR without producing a PCR product. Primer 7, MRF/AsyR, works as the Allele specific forward primer to the AsyR primer. In the primers, blue indicates the SNP nucleotide base position in the primer sequence. The red denotes the deliberately introduced mismatches to increase the specificity.

6.1.4 Two-way allele specific PCR

I wanted to increase the stringency of ASPCR primers. Therefore, I designed both the forward and reverse primers to bind at the SNP site at their 3' end, and named it the two-way ASPCR test (Figure 6.5). This strategy might work best for the allele specific PCR targets that do not have the pseudogenes to cross-contaminate the PCR reaction by offering a potential binding site for one of the actual target's primers. The primers were ordered before I started troubleshooting the problem of pseudogenes. Primers 1-4 in Table 6.4 were designed to perform 2 way allele specific PCR, where both forward and reverse primer 3' ends match with the SNP, and they form a very short product of 50 base pairs (25 bp of forward and 25 bp of reverse primer). Primer 5 in Table 6.4 WR43B has a deliberately introduced mismatch at the antepenultimate (-2) position to see its effect on ASPCR. However, it does not show allele specific nature with any of the above forward primers. Thus use of these primers was abandoned.

Figure 6.5 Two-way Allele Specific PCR



The 3' binding site of the primers is at the SNP base. When the primers match (1), both the primers were extended during PCR to make the product. When they mismatch at their 3' binding site (2), the PCR does not occur. This method makes the general ASPCR approach more stringent. Here circle in the solid line is the SNP base. Melt profile of the PCR product is only obtained if there is a match between the primers and SNP and the PCR happens.

Table 6.4 Two-way allele specific PCR

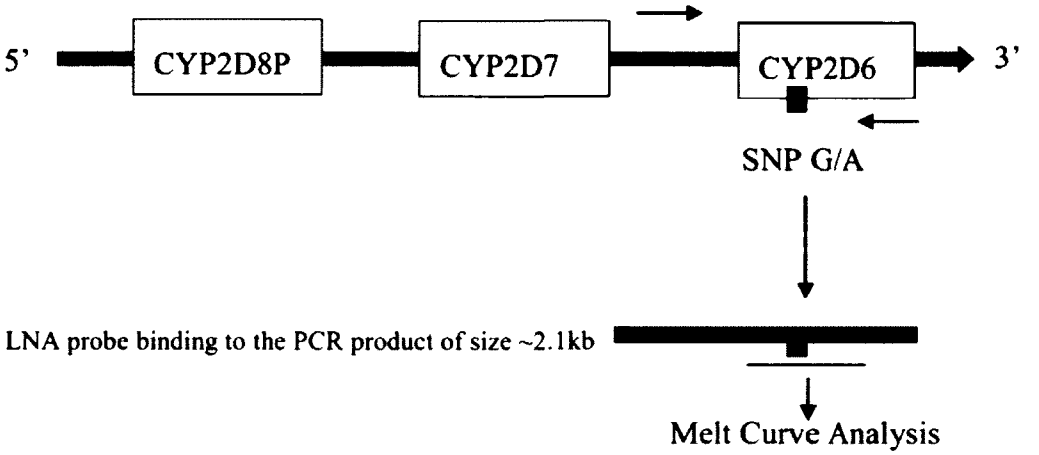
S.No	Primer	Primer Sequence	Product Size
2 way allele specific PCR primers			
1	AS-R25	5'gaccgttggggcgaaaggggcgtcc	50bp with #3-4
2	AS-R252A	5'gaccgttggggcgaaaggggcgtac	
3	AS-F25	5'ccttacccgcatctcccacccccag	50bp with #1-2
4	AS-F252T	5'ccttacccgcatctcccacccccctg	
WR4 further changed			
5	WR4-3B	5'cgttggggcgaaaggggcgacc	

Legend to table 6.4. Both the primers' 3' end bind on the SNP position. These primers were only meant for allele specific PCR. They do not distinguish between active and pseudogenes. Red denotes the deliberate mismatches introduced in the primers to increase the specificity, whereas the blue base indicates the SNP position in the primer.

6.1.5 Asymmetric PCR with Locked Nucleic Acid (LNA) probe

The proposed ASPCR approach that excludes the pseudogenes was producing secondary products. To solve this problem, I proposed to amplify the SNP carrying sequence by employing the forward primer specific to the intronic region of *CYP2D6* gene and the other primer binding a few bases upstream to the SNP site, creating a 2140bp PCR product (Table 6.5). Amplification of the gene was followed by detection of the SNP (Figure 6.6). To detect the SNP, an LNA probe was designed to bind at the SNP site. The probe binds to the PCR product of 2140bp size at the SNP location, and the probe was melted during melt curve analysis allowing discrimination between wild and mutant SNP with a melt temperature difference of 7-8°C between the wild and mutant alleles.

Figure 6.6 LNA probe binding to the PCR product



The red color is the SNP base that binds at the center of the probe. Probe is matched to the wild type and melts at different temperature compared to when it mismatches with a mutant SNP allele. However, the large PCR product had several double stranded loops that interfered with the probe melt peaks during the MCA.

Table 6.5 LNA probe based SNP genotyping

S.No	Primer	Primer Sequence	Product Size
1	P100	5' ggcctaccctgggtaagggcctggagcagga	2140 bp
2	Seq64R	5' agaccgttggggcgaaaggg	
3	LNA Probe CYP2D6*4-G1	5'cgtcctggg/3spc3/	9 bp

After the target region was amplified, the LNA probe binds to the SNP region. The probe was complementary to wild type allele 'G'. Therefore, the probe has a higher Tm difference between wild and mutant allele when compared to an unmodified probe.

In practice, this concept could have generated good results if the PCR product was within ~500bp. However, the 3kb duplex product did not melt at a single temperature. Instead, the 3kb product was coiled into multiple DNA-DNA duplex loops, and melts at various temperatures raising the background noise. Therefore, it was very difficult to differentiate the probe-DNA duplex melting from the background noise. Therefore, this method was abandoned.

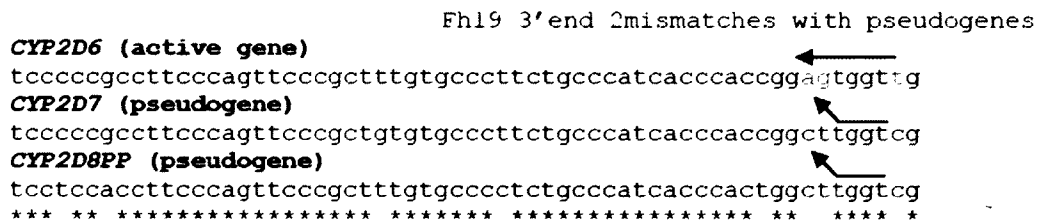
6.1.6 Outcome of the primer designs

The above primer sets were tested and most were abandoned for reasons of specificity involving pseudogenes. However it was important to show the range of combinations that were designed and tested. Primers with more than one deliberate mismatch prior to the 3' end did not work. The primer with one deliberate mismatch gives a PCR product. Among the allele specific primers, the Fh19 (Table 6.2) and WR42G (Table 6.3) primer set gave better PCR product yield. The primer Fh19, as stated earlier, discriminate the pseudogene by binding the active gene variation sites at its 3' end. To provide the proof of concept, multiple sequence alignment of pseudogenes and active gene, and the *CYP2D6* sequence specific position where Fh19 binds are shown in clustal format using MAFFT V6.833b (Figure 6.7, a multiple sequence alignment tool). When the BLAST was performed for the primer pair, they do not show amplification of pseudogenes, but they appear to be specific to *CYP2D6* only. The forward primer Fh19 and the reverse primer WR42G were not allele specific to some of the mismatched samples in the conventional PCR and gave the PCR product. Similarly, in the gel post system the primers did not show allele specificity to few

mismatched samples. It proves that the primers do not make an efficient primer pair to perform ASPCR of the active gene *CYP2D6* in the presence of the pseudogenes *CYP2D7* and *CYP2D8P*. The explanation for not producing an allele specific PCR is cross binding of Fh19 to pseudogenes. However, no reports have been found that verify the frequency of penetrance for these variation sites, or confirm the consistency of these variations in the Caucasian population.

There are numerous variations across the complete pseudogene sequence. However, there is only one mutation reported by Kimura et al [1], being consistent in the Caucasian population. The mutation is the insertion of T at base upstream +226 in the first exon that disrupts the protein reading frame, thereby preventing the conversion of mRNA to a functional P450 enzyme. The solution to the pseudogene problem was to design a primer that avoids pseudogene amplification by binding at its 3' end to the mutation site.

Figure 6.7 Considering variation site to avoid pseudogene amplification by common forward primer (Fh19) in allele specific PCR.



CLUSTAL format alignment by MAFFT (v6.833b). Red color AG bases are the sequence where Fh19 binds at its 3' end. Mismatches between Fh19 and pseudogenes are in a rectangular block and the mismatch bases in pseudogenes are in blue.

6.2 Guidelines for ASPCR

6.2.1 Allele Specific Primer characteristics

1) Primer length should be within 18-25 bases long.

2) The 3' end of the primer should not contain GC repeats.

3) Check for intra-repeats within primer to make sure no hairpins are formed

4) Check for inter repeats between two primers to make sure no primer duplex are formed.

5) 3' end of the primer should not end in T; otherwise it loses allele specific nature.

6) GC content of the primer to be around 40-60%.

7) Other characteristics like low dNTP concentration, low primer concentration, low $Mgcl_2$ concentration, low enzyme concentration, high annealing temperature and limiting the PCR cycles will increase the specificity or stringency of the reaction. However, there will be a decrease in overall PCR product yield of the reaction. PCR conditions have to be optimized to create a balance between the product yield of the reaction (producing detectable PCR product) and specificity (distinguishing between match and mismatch alleles).

6.2.2 Enzymes for ASPCR [2]

The following enzymes do not have 3'-5' exonuclease activity, which is essential to perform ASPCR, as exonuclease activity corrects the introduced mismatches in the primers [3].

The enzymes are compared to *Taq* polymerase for the PCR yield and the ASPCR stringency. However, the PCR yield of the reaction mostly depends upon 1) the primer design 2) how well the reaction is optimized to the conditions and reagent concentrations.

- Tbr, Stoffel fragment of *Taq* polymerase – better or equal to *Taq* polymerase
- Hot Tub, Pfu, Pfu[exo-], Ultma almost similar to *Taq* polymerase

6.2.3 Guidelines for introducing a deliberate mismatch at 3'end of the AS primer [4-8]

The 3' end of the allele specific primer sits at the SNP location, and serves as mismatch for non-complementary SNP alleles. Additional mismatches are introduced at the penultimate (-1) or anti-penultimate (-2) position at the 3' end to increase the stringency of the allele specificity in the reaction.

Note: The first base in the mismatch pair belongs to the primer and the later to the template.

- A:G, G:A, A:A, C:C (purines: purines) combination mismatches are strong with the exception of C:C & G:G. (20-100 fold reduction in extension/ NO extension). However, C:C & G:G can show moderate extension

- AC, CA, GT mismatches show moderate extension, they can still be incorporated depending upon the strength of mismatch at the 3' end (complementary to the SNP base).

- TG, AC, CA, mismatches are more likely to extend

- AT, AG, TT, TC, TA, CT, GC, CG mismatches do not likely discriminate the SNP allele.

Overall, the mismatches that are in decreasing order of ASPCR stringency are: AA>CC>AG>GG>TT>AC, GT>GA. Therefore, in general, purine-purine mismatches add more stringency to allele specificity compared to purine-pyrimidine or pyramidine-pyramidine mismatches.

6.2.4 Optimization of ASPCR

1) Magnesium Concentration: Concentrations below 1.5mM can increase the specificity of the reaction. Concentrations above 4mM can produce non-specific products.

2) Primer Concentration: Lowering the concentration of the primer from 0.2μM-0.05μM can increase the specificity. However, primer

concentration below 0.025μM will produce weak products, and it is not preferred.

3) DNA concentration: Ten-fold dilution of standard pDNA concentration or genomic DNA concentration (2.5ng/μl) can still give an adequate product and increase the specificity of the reaction.

4) Product size: Allele specific PCR can be obtained for 150 to ~2kb products. However, the favourable range of PCR product is 200-500bp.

5) Dexoynucleotide concentration: Lowering the dNTP concentration from 200μM to 50μM can increase the specificity of the reaction and reduce the non-specific product formation.

6) Formamide: 2-5% formamide can minimise non-specific products from a G-C rich target sequence.

7) *Taq* Polymerase: The *Taq* should not have 3'-5' exonuclease activity. Lowering the concentration of *Taq* from 5 U/μl to 0.5μl can increase the specificity of the reaction, and helps in reducing the formation of non-specific products.

8) PCR cycles: Lowering the number of cycles will help in reducing the mismatch product and primer dimer extension product. In general, 30-35 cycles are preferred.

9) PCR thermal conditions: Initial denaturation temperature of 94°C for one minute, followed by denaturation at 94°C for 15-40 seconds, annealing temperatures 50°C -72°C for 30 seconds - 1 minute, and extension of 72°C for 30 seconds – 2 minutes, and final extension of 72°C for 2-6 minutes are sufficient. The longer duration and higher temperatures are favourable for highly GC rich target region or the primers with higher melting temperatures (Tm value).

10) Primer placement: Each allele specific primer (wild or mutant type) can be paired with a common forward primer in each tube (in total 2 tubes in total). However, if the specificity criteria is not met, the common forward primer can be replaced by an allele specific primer, and it becomes a two-way allele specific PCR. To increase the product size in such cases, a few thymine bases can be added to its 5' end. Alternatively a number of deliberate mismatches can be introduced at the 3' end of the primer [9], extra care has to be taken as it significantly reduces the PCR yield.

11) Tm values for the allele specific primers: The Tm values of common forward, allele specific reverse (wild and mutant) should have a close value (1°C -3°C). It helps in performing both the ASPCR reactions (wild and mutant) simultaneously on single thermocycler.

12) ASPCR with heterozygous sample: Heterozygous samples are statistically and practically higher in proportion in the population when

compared to homozygous samples (wild or mutant). In the heterozygous samples, the available concentration of match DNA is reduced by half, since it has both wild and mutant alleles in equal proportion. Extra care has to be taken in optimizing the number of PCR cycles which should be enough to extend the PCR product in both heterozygous and homozygous match condition [10].

6.2.5 Online primer designing tools

Some of the important primer designing tools include:

1) WASP for ASPCR primer design along with penultimate base change [11]. http://bioinfo.biotec.or.th/WASP

2) Autodimer Tool assesses primer dimmer and hairpin formation.

3) PRISE (PRImer SElector) [12].

   http://www.jcsg.org/prod/scripts/primer/primer_input_form.cgi

4) Auto Prime designs primer for real-time PCR measurement of eukaryotic gene expression.

5) CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide) Primers are designed from protein multiple sequence alignments.

6) IDT Anti Sense Design for antisense oligo design and selection tool.

7) IDT Oligo Analyzer Online calculates oligonucleotide parameters such as melting temperature. Shows self-dimers, hairpin, and performs Blast.

8) IDT Primer Quest Primer and probe design and selection.

9) Net Primer Java applet for primer design.

10) Oligonucleotide Analyzer generates Tm, free energy, molecular weight and hairpin and dimer formation structures.

11) Primer3 Utility locates oligonucleotide primers for PCR amplification of DNA sequences.

12) UCSC In-Silico PCR In-Silico PCR searches a genome sequence database with a given pair of PCR primers.

6.2.6 Steps involved in designing an optimum primer pair

1) Find SNPs using SNPedia, NCBI or other links available in SNPedia.

2) Go to gene window (online tool) or Mapveiwer in NCBI to see the whole gene sequences and the sequence flanking to SNP.

3) Copy and paste the sequence in the WASP online tool for designing ASPCR primers.

4) Consider all the guidelines for making primers (Length, GC content, Tm, hairpin or primer dimer formations).

5) Make sure the primers and the product region do not have another major SNP.

6) Use Autodimer to see hairpin and primer dimmer formation.

7) Introduce more deliberate mismatches if required.

8) Use more than 3 online primer tools and paste the primer sequence to analyze overall quality of the primer set.

9) BLAST the primer sequence to see if it amplifies any of the pseudogene sequences / secondary products.

References

[1] S. Kimura, M. Umeno, R. C. Skoda, U. A. Meyer and F. J. Gonzalez, "The human debrisoquine 4-hydroxylase (CYP2D) locus: Sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene," *Am. J. Hum. Genet.*, vol. 45, pp. 889-904, 1989.

[2] J. M. Gale and G. B. Tafoya, "Evaluation of 15 polymerases and phosphorothioate primer modification for detection of UV-induced C:G to T:A mutations by allele-specific PCR," *Photochem. Photobiol.*, vol. 79, pp. 461-469, 2004.

[3] S. Ayyadevara, J. J. Thaden and R. J. Shmookler Reis, "Discrimination of primer 3'-nucleotide mismatch by *Taq* DNA polymerase during polymerase chain reaction," *Anal. Biochem.*, vol. 284, pp. 11-18, 2000.

[4] S. Kwok, D. E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson and J. J. Sninsky, "Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies," *Nucleic Acids Res.*, vol. 18, pp. 999-1005, 1990.

[5] S. Kwok, S. Chang, J. J. Sninsky and A. Wang, "A guide to the design and use of mismatched and degenerate primers," *Genome Res.*, vol. 3, pp. S39-S47, 1994.

[6] Y. Wang, J. A. Rollin and Y. Percival Zhang, "Enhancing allele-specific PCR for specifically detecting short deletion and insertion DNA mutations," *Mol. Cell. Probes,* vol. 24, pp. 15-19, 2010.

[7] C. R. Newton, A. Graham, L. E. Heptinstall, S. J. Powell, C. Summers, N. Kalsheker, J. C. Smith and A. F. Markham, "Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)," *Nucleic Acids Res.*, vol. 17, pp. 2503-2516, 1989.

[8] P. Kwok, "Methods for genotyping single nucleotide polymorphisms," *Annual Review of Genomics and Human Genetics,* vol. 2, pp. 235-258, 2001.

[9] M. Huang, N. Arnheim and M. F. Goodman, "Extension of base mispairs by *Taq* DNA polymerase: Implications for single nucleotide discrimination in PCR," *Nucleic Acids Res.*, vol. 20, pp. 4567-4573, 1992.

[10] P. S. Walsh, H. A. Erlich and R. Higuchi, "Preferential PCR amplification of alleles: Mechanisms and solutions," *Genome Res.*, vol. 1, pp. 241-250, 1992.

[11] P. Wangkumhang, K. Chaichoompu, C. Ngamphiw, U. Ruangrit, J. Chanprasert, A. Assawamakin and S. Tongsima, "WASP: A Web-based Allele-

Specific PCR assay designing tool for detecting SNPs and mutations," *BMC Genomics,* vol. 8, 2007.

[12] Q. Fu, P. Ruegger, E. Bent, M. Chrobak and J. Borneman, "PRISE (PRImer SElector): Software for designing sequence-selective PCR primers," *J. Microbiol. Methods,* vol. 72, pp. 263-267, 2008.