Microarray and Molecular Genetic Analysis of Aberrant Splicing in Human Drug Metabolizing Cytochromes P450 CYP2D6 and CYP2B6

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart
zur Erlangung der Würde eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

Vorgelegt von

Marco Hans Hofmann

aus Weiden i.d.Opf

Hauptberichter: Prof. Dr. R. D. Schmid
Mitberichter: Prof. Dr. F. Wollnik
Mitberichter: Prof. Dr. U. M. Zanger

Tag der mündlichen Prüfung: 24. Juli 2008

Dr. Margarete Fischer-Bosch Institut für Klinische Pharmakologie, Stuttgart
und
Institut für Technische Biochemie der Universität Stuttgart
2008

Jeder Strom braucht eine starke Quelle und beständige Nebenflüsse
Für meine Eltern und meine Freunde

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Teile der Arbeit wurden bereits veröffentlicht in:

- U.M. Zanger and M.H. Hofmann, Polymorphic Cytochromes P450 *CYP2B6* and *CYP2D6*: Recent Advances on Single Nucleotide Polymorphisms Affecting Splicing. Acta Chim. Slov. 2008, 55(1), 38-44

- M.H. Hofmann, J.K. Blievernicht, K. Klein, T. Saussele, E. Schaeffeler, M.Schwab, U.M. Zanger, Aberrant Splicing Caused by Single Nucleotide Polymorphism c.516G>T [Q172H], a Marker of *CYP2B6*6*, Is Responsible for Decreased Expression and Activity of CYP2B6 in Liver. J Pharmacol Exp Ther. 2008, 325(1), 284-92

List of Abbreviations

ADR, adverse drug reactions

APS, ammoniumpersulfate

bp, base pair

BSA, bovine serum albumin

cDNA, complementary DNA

cRNA, complementary RNA

CNS, central nervous system

CYP, cytochrome P450

DMEM, Dulbecco's modified eagle medium

DMSO, dimethylsulfoxide

dNTP, deoxyribonucleotide triphosphate

DNA, deoxyribonucleic acid

EDTA, ethylendiamintetraacetic acid

EFV, efavirenz

EM, extensive metabolizer

EMSA, electrophoretic mobility shift assay

ESE, exonic splicing enhancer

ESS, exonic splicing silencer

HIV, human immunodeficiency virus

IM, intermediate metabolizer

ISE, intronic splicing enhancer

ISS, intronic splicing silencer

MALDI-TOF MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry

MEM, minimum essential medium

NMD, nonsense-mediated decay

NP, normal product

Nt, nucleotide

NT/F, number of nucleotides per fluorophore

PCR, polymerase chain reaction

PM, poor metabolizer

PTC, premature termination codon

RNA, ribonucleic acid

RT, reverse transcription

SAP, shrimp alkaline phosphatase

siRNA, small interfering RNA

SDS, sodiumdodecylsulfate

SNP, single nucleotide polymorphism

snRNA, small nuclear RNA

SR-protein, serine-arginine protein

SV, splicing variant

Tm, melting temperature

UM, ultrarapid metabolizer

β-gal, β-galactosidase

ΔG, enthalpy

Zusammenfassung

Die Superfamilie der hochpolymorphen Cytochrome P450 Enzyme besteht aus 57 funktionellen Genen und 58 nicht funktionellen Pseudogenen. Diese Doktorarbeit widmet sich der Detektion der häufigsten Spleißvarianten in den Cytochrom P450 Enzymen CYP2D6 und CYP2B6 und der Fragestellung, ob diese mechanistisch bestimmten Punktmutationen zugeordnet werden können.

Für die Untersuchung der mRNA-Transkripte in beiden Genen ist ein Mikroarray entwickelt worden, der hoch sensitiv alternative Spleißereignisse detektiert. Der Mikroarray wurde in allen Schritten optimiert, angefangen von der Auswahl der Oligonukleotidsonden, des Herstellungsprozesses der Arrayplattform, dem Umschreiben und Markieren der Target-DNA, den Hybridisierungsbedingungen als auch der Entwicklung einer neuen Auswertemethode, die mit dem gewählten Design der Sonden verwendet werden kann. Um unterschiedlichste Spleißereignisse detektieren zu können, wurde ein Design gewählt bestehend aus fünf verschiedenen Sonden pro Exon: Exon-Sonden, Intron-Exon-, Exon-Intron-, Exon(n)-Exon(n+1) (junction) und Exon(n)-Exon(n+2) (jump junction) überlappende Sonden. Alle Sonden wurden zunächst mit bekannten DNA-Fragmenten validiert. Es steht am Dr. Margarete Fischer-Bosch Institute eine umfangreiche und gut dokumentierte geno- und phenotypisierte humane Lebersammlung zur Verfügung. Das Spleißingmuster von ausgesuchten Proben der Sammlung wurde mit Hilfe des entwickelten Mikroarrays untersucht.

Es ist bekannt, dass in Allel *CYP2D6*41* die Mutation 2988G>A in Intron 6 das Spleißmuster zu einer Transkriptvariante verschiebt, in der Exon 6 fehlt. Der entwickelte Mikroarray für *CYP2D6* bestätigte das gehäufte Auftreten dieser alternativen Spleißvariante in Allel *CYP2D6*41* und bewies damit seine Funktionalität. Mit Hilfe des *CYP2D6*-Microarrays wurden zudem allelspezifische Spleißmuster bei den Trägern der häufigsten *CYP2D6*-Allele **1*, **2*, **4* und **41*, bestimmt. Es konnte beobachtet werden, dass Träger des Allels *CYP2D6*41,* zusätzlich zu dem bekannten alternativen mRNA-Transkript, in denen Exon 6 fehlt, vermehrt Transkripte besitzen, in welchen Intron 5 und 6 nicht herausgespleißt worden sind. Die Transkriptverteilung in Allel **1* und **2* ist vergleichbar. Beide besitzen fünf Mal mehr des funktionellen Transkriptes (NP), verglichen mit **41*.

Somit erwiesen sich die entwickelten Spleißarrays als nützliches Werkzeug, um nicht nur Lebergewebe sondern auch allelspezifische Spleißmuster zu detektieren.

Die Existenz von nicht funktionellen Pseudogenen mit hoher Homologie zu den eigentlichen funktionellen Genen, der Cytochrom P450 Enzyme, stellt bei der Interpretation von Daten, die von konventionellen Spleißarrays stammen, ein großes, bisher wenig beachtetes Problem dar. Durch das spezifische Sondendesign, in Kombination mit gen- bzw. pseudogenspezifischer reverser und linearer Amplifikation, ermöglicht die entwickelte Arrayplattform erstmals, sowohl Gen als auch pseudogenspezifisches Spleißen zu detektieren und zu unterscheiden. Verglichen zum funktionellen Gen *CYP2D6* beinhalten die mRNA Transkripte des Pseudogen *CYP2D7* häufiger Intron 5 und 6. Im Vergleich zu *CYP2D6* wurden in *CYP2D7* keine erhöhten Signale detektiert, die auf ein Fehlen von Exon 6 schließen lassen.

Am Beispiel des *CYP2B6* Gens konnte gezeigt werden, dass sich die entwickelte Mikroarrayplattform für *CYP2D6* und *CYP2D7* auch auf andere Gene übertragen lässt. Alternatives Spleißen wurde in *CYP2B6* bisher nicht ausführlich untersucht. Das Genprodukt von *CYP2B6* ist ein hochpolymorphes Enzym, welches eine wichtige klinische Rolle bei dem Verstoffwechseln einer großen Anzahl von Arzneimittel, wie z.B. Cyclophosphamid, Bupropion und Efavirenz, spielt. Das häufige Allel *CYP2B6*6* [c.516G>T, Q172H and c.785A>G, K262R] zeigt erniedrigte Proteinexpression in humaner Leber und wurde erst kürzlich mit erhöhtem Efavirenz-Plasmaspiegel in HIV Patienten in Verbindung gebracht. Der molekulare Mechanismus für diese Beobachtungen war bis jetzt ungeklärt.

Mit dem entwickelten Spleißarray für *CYP2B6* wurde eine neue Spleißvariante gefunden, in der Exon 4 fehlt (SV8) sowie auffallende Transkriptunterschiede beim Vergleich von *CYP2B6*6* und **1*. Dies führte zu der Idee, dass alternatives Spleißen im Allel **6* eine Rolle spielen könnte. Deswegen wurde RNA aus gut dokumentierten humanen Lebergeweben molekularbiologisch untersucht. Es zeigte sich, dass bisher unbekannte Spleißvarianten (SV8, SV7, SV9) existieren, die sich aber alle als genotypunabhängig herausstellten. Das häufigste Transkript in *CYP2B6*6* war jedoch nicht das normale mRNA-Transkript (NP) mit allen neun Exons, sondern ein alternatives Transkript, in dem Exon 4 bis 6 fehlt (SV1). SV1 trat in Leberproben mit dem Allel **6* verstärkt auf und war zudem im Zusammenhang mit dem seltenen Allel *CYP2B6*3* (c.777C>A) verstärkt nachweisbar. Diese Beobachtungen an humanen

Leberproben weisen auf alternatives Spleißen als mögliche Ursache für erniedrigte Genexpression bei varianten *CYP2B6* Allelen hin.

Weitere Untersuchungen mit Hilfe von Transfektionsexperimenten in COS-1 und Huh7 Zellen wurden mit Hilfe eines in dieser Arbeit klonierten *CYP2B6*-Minigenkonstrukt durchgeführt. Das klonierte Minigenkonstrukt von *CYP2B6* enthält alle neun Exons sowie fünf ganze oder teilweise verkürzte Introns. Durch das Einfügen von Punktmutationen in dieses Konstrukt ist es möglich, alternatives Spleißen in Abhängigkeit von Einzelmutationen in Zellsystemen zu studieren.
Mit Hilfe der Minigenkonstrukte zeigte sich deutlich, dass allein der Nukleotidaustausch c.516G>T in Allel *CYP2B6*6* verantwortlich ist für das verschobene Spleißmuster zugunsten des Transkripts SV1 und einer verringerten *CYP2B6*-Expression. Minigenkonstrukte mit dem alleinigen Nukleotidaustausch c.785A>G oder der seltenen Mutation c.777C>A führen dagegen zu normaler bzw. mittlerer Expression des Enzyms CYP2B6.
Zusammenfassend kann gesagt werden, dass der Mechanismus des häufigen Allels**6* hauptsächlich auf einem prätranslationalen Mechanismus, dem alternativen Spleißen, als Folge einer nicht-synonymen, exonischen Mutation beruht.

Diese Arbeit zeigt die Auswirkung von Punktmutationen in Spleißmotiven, als auch die klinische Bedeutung von alternativem Spleißen bei der Aufklärung von medikamentösen Nebenwirkungen wie im Beispiel von Efavirenz in *CYP2B6*6*-Trägern. Sowohl in *CYP2D6* als auch *CYP2B6* existiert somit je ein häufiges Allel (*CYP2D6*41* bzw. *CYP2B6*6*), in welchem durch alternatives Spleißen die Menge an funktioneller mRNA und damit Protein und Enzymaktivität verringert wird. Die damit gesicherte Erkenntnis der verursachenden Mutation stellt somit eine sichere Grundlage dar für die Anwendung dieses Polymorphismus in der genetischen Diagnostik sowie der klinischen Praxis.

Abstract

In humans the polymorphic Cytochrome P450 gene superfamily consist of 57 functional genes and 58 highly homologous pseudogenes. This study was devoted to the detection of alternative splicing within the Cytochrome P450 enzymes 2D6 and 2B6, mapping of the most common splice variants and to draw connections to certain single nucleotide polymorphisms (SNPs) and alleles. For both enzymes a splicing sensitive microarray was developed. The microarray was produced and optimized in all steps including the oligonucleotide probe design, microarray processing and target preparation, optimization of hybridization conditions and the development of a new data quantification method for the used probe design.

For the developed splicing platform a design was chosen based on 5 different probes: exon-, intron-exon-, exon-intron-, junction- and jump-junction-probes to detect the most common splicing events. All probes were evaluated with known DNA fragments. In the Dr. Margarete Fischer-Bosch Institute there exists a collection of well-characterized pheno- and genotyped human liver samples from patients. Selected samples from this collection were chosen for investigation of alternative splicing with the developed splicing microarray.

Within the *CYP2D6* gene it was known that the SNP 2988G>A (allele **41*) in intron 6 shifts splicing towards a variant lacking exon 6, what explains the intermediate phenotype within allele **41*. The splicing platform verified this splicing aberration in allele **41*. Using the microarry specific splicing patterns were monitored in human liver tissue within the most common alleles of *CYP2D6 *1*, **2*, **4* and **41*. It could be observed that within mRNA from allele **41* carriers additionally to the known transcript variant, which is lacking exon 6, total or partial retention of intron 5 and 6 was enhanced. Transcript patterns of *CYP2D6*1* and *2 were similar with 5 times higher amount of the full functional transcript (NP), including all nine exons, compared to allele **41*. The splicing array showed to be a valuable tool not only for detection of splicing variants in human liver tissue but additionally for detection for allele specific splicing patterns.

The existence of highly homologous Cytochrome P450 pseudogenes, which in some cases, as in *CYP2D7* also express alternative splicing variants, results in a major problem of interpreting the data from splicing arrays. The developed splicing platform

is the first existing array with which gene and pseudogene specific transcript patterns can be monitored individually. This was achieved by using gene and pseudogene specific probes in combination with a gene specific reverse and linear amplification protocol of the target RNA. Within *CYP2D7* mRNA a higher amount of full or partial retention of intron 5 and 6 occurred independent of any *CYP2D6* genotype. Transcripts lacking exon 6 were not observed within *CYP2D7*.

The microarray platform can be easily transferred to other genes as shown for the second gene *CYP2B6*. Alternative splicing in this gene was so far only reported descriptive. *CYP2B6* is a polymorphic human drug metabolizing cytochrome P450 with clinical relevance for several drug substrates including cyclophosphamide, bupropion and efavirenz. The common allele *CYP2B6*6* [c. 516G>T, Q172H and c.785A>G, K262R] has previously been associated with lower expression in human liver and with increased plasma levels of efavirenz in HIV patients, but the molecular mechanism has remained unclear. With the developed splicing array for *CYP2B6* allele specific splicing patterns were observed comparing *CYP2B6*6* and *CYP2B6*1*. This lead to the idea that alternative splicing might play an important role in allele **6.* This was investigated in more detail using RNA originating from well-documented human liver tissue. Analysis of mRNA in this tissue demonstrated that additional unknown splicing variants exist (SV8, SV7, SV9), which all were genotype independent.

Investigations in human liver tissue using RT-PCR and sequencing showed that the most common transcript in *CYP2B6*6* was not the normal transcript (NP) but an alternative splicing transcript lacking exons 4 to 6 (SV1). SV1 was tightly associated with the allele**6* and apparently also with the rare variant c.777C>A (*CYP2B6*3*). The observations lead to the assumptions that alternative splicing might explain the decreased function observed in allele *CYP2B6*6*.

Further investigations in this direction were performed by cloning *CYP2B6* minigene constructs including all nine exons and additional intronic regions. With site directed mutagenesis common mutations were added into the minigene. This enables the investigation of alternative splicing of *CYP2B6* in cell systems and to draw connections to certain SNPs. The transfection of the *CYP2B6* minigene in eukaryotic cell lines COS-1 and Huh7 demonstrated that the single nucleotide polymorphism c.516G>T in allele *CYP2B6*6* was alone responsible for aberrant splicing resulting in high SV1 and low CYP2B6 expression phenotype. Minigenes carrying the single

c.785A>G polymorphism or the rare c.777C>A variant resulted in normal and intermediate expression phenotypes, respectively. In conclusion, the mechanism of the common allele**6* involves predominantly a pretranslational mechanism resulting in decreased enzyme expression. Aberrant splicing is leading to reduce functional mRNA, protein and activity. These results establish the SNP c.516G>T, a nonsynonymous exonic mutation, as the causal sequence variation for severely decreased expression and function associated with *CYP2B6*6*.

This work emphasizes the role of SNPs in non-consensus splicing elements such as exonic and intronic splicing enhancers as well as the clinical relevance of alternative splicing in context of adverse drug reactions. In both investigated genes *CYP2D6* as well as in *CYP2B6* there exists a common allele (*CYP2D6*41* and *CYP2B6*6*, respectively*)* in which aberrant splicing results in reduced amounts of functional transcript, reduced amount of protein and enzyme activity. The findings establishes the SNP c.516G>T as the causal sequence variation that can now be reliably used in pharmacogenetic studies in various clinical settings including prediction of drug plasma concentration, toxicity, drug effectiveness and dose adjustment.

1 Introduction

The human genome project started in 1990 with the first draft being completed by 2003 and completed at high confidence by 2006, when the sequence of the last chromosome was published in Nature (Gregory *et al.*, 2006). Very astonishing at this time, one major result was that within the 3 billion base pairs in the human genome only around 22000-23000 genes were found, about half the number previously predicted by many scientists (Gregory et al., 2006; Hattori, 2005). Compared to the genome of *Drosophila melongaster* with about 13000 or the small roundworm (*Caenorhabditis elegans*) with 19000 genes (C.elegans Sequencing Consortium, 1989; Gregory et al., 2006; Hodgkin, 2001; Myers et al., 2000) the obvious question arose, how the complexity of the human organism could be explained with such a comparatively small number of genes.

In principal, the higher organisation of humans or mammals could be explained by further complexity of genetic organisation, for example by modification of the DNA and organisation of chromatin structure, within complex regulation of transcription including subsequent splicing and finally translational and posttranslational modification. Biological complexity would depend less on gene number but on the transcripts, which are expressed largely due to epigenetic mechanisms.

1.1 Complexity of gene regulation

1.1.1 Modification of DNA

Modifications within the DNA are mainly due to epigenetic mechanisms. “Epigenetic“ (coined by C.H. Waddington in 1942) describes the study of heritable changes in gene expression or function that occur without changes in the DNA sequence itself (Waddington, 1959). Several types of epigenetic mechanisms play a role on the DNA level as part of the biological organisation thus increasing complexity. There are three main types of epigenetic mechanisms on the DNA level: cytosine DNA methylation, histone modifications and genome imprinting.

With genome imprinting the diploid euchromatic chromosome gets heterochromatic so that imprinted genes are only expressed from one parental allele. This mechanism is based on methylation and histone modification so more or less comprising the two other mechanisms (Wood and Oakey, 2006). Within the nucleus DNA does not exist naked but wrapped around histone octameres. Histone modification shows modulation effects on gene expression or gene silencing and can be altered by acetylation, phosphorylation, ubiquitination as well as methylation in accessible and inaccessible regions. This has direct effects on the transcriptional regulation (He and Lehming, 2003). DNA methylation on so-called CpG islands within the promoter region usually silences gene expression. It is known that disease can be associated with a different methylation pattern. For example the cancer epigenome differs from the genome by specific methylation pattern. Changes are encoded by genes like DNA methyltransferases (DNMT), methyl-CpG-binding domain (MBD) proteins, histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases (HMT) and histone demethylases (Esteller, 2007; Miremadi et al., 2007).

1.1.2 Regulation of transcription

Transcription is the mechanism by which RNA is synthesised by RNA polymerase using DNA as template. RNA synthesis was first discovered in 1962 by Yanofsky (1962) and Giacomoni (1964). Transcription can be regulated by specific transcription factors and cofactors that alter the specificity of the RNA polymerase for a given promoter or a transcription start site. Repressors bind to sequences next to

the promoter and this way prevent formation of the initiation complex. Enhancers are sites on the DNA on which activators, the opponents to the repressors, aid to build the initiation complex or change the secondary structure of the DNA so that transcription is supported. Within an inducible system no gene expression takes place unless a special molecule called inducer is presente or binds to the DNA. The induction of many CYPs including *CYP3A4*, *CYP2B6* as well as *CYP1A1* occurs by ligand activation of key receptor transcription factors including the pregnant X-receptor (PXR), the constitutive androstane receptor (CAR) and the aryl hydrocarbon receptor (AhR) (Faucette et al., 2006; Faucette et al., 2007; Smith et al., 2005; Tompkins and Wallace, 2007; Wang et al., 2004). Within a repressible system expression takes place except a molecule is present or binds to the DNA. This mechanism is less frequent in cytochrome P450 family but is described for example in *CYP27B1* which expression is negatively regulated via vitamin D response element (Turunen *et al.*, 2007).

Mutations within the promoter region or within regulating elements can alter the binding of inducers, repressors or formation of the transcription complex at the start side and so influence the expression. Additionally multiple transcription initiation complexes for one gene are possibly increasing protein diversity (Quelle *et al.*, 1995).

1.1.3 Posttranscriptional modification

Posttranscriptional mechanisms that increase protein diversity include capping, polyadenylation, pre-mRNA editing (Maas *et al.*, 2006), and alternative splicing. As soon as transcription started at the 5'end the capping enzyme complex bound to RNA polymerase begins the capping process. A guanine nucleotide is connected via an unusual 5' to 5' triphosphate linkage and will be methylated on the 7' position directly after capping by a methyl transferase (Blencowe, 2006). Transcription termination in eukaryotes involves cleavage of the new transcript 10 - 15 nucleotides behind a polyadenylation sequence AAUAAA and addition of 50 -150 adenosines at the 3'tail. 5'capping as well as 3'polyadenylation is necessary for export from the nucleus, provides significant resistance to 5'exonucleases and specially the 5'cap structure initiates the 5'proximal intron excision (Colgan and Manley, 1997; Lewis et al., 1995; Wahle and Ruegsegger, 1999). Furthermore pre-mRNA adenosine to inosine editing catalysed by adenosine deaminases can entail further changes in protein structure and function (Maas et al., 2006).

All those described posttranscriptional mechanisms alter and influence pre-mRNA splicing which is considered to be the most important source of protein diversity in vertebrates with 40-70% of all human genes having alternative splicing transcripts (Graveley, 2001; Johnson et al., 2003; Modrek and Lee, 2002; Modrek et al., 2001).

1.2 Splicing

The mechanism of splicing is used to remove intron regions from pre-mRNA and junction exon regions. The mean human exon is 150 bp, the average intron is 3500 bp long (Deutsch and Long, 1999). The mechanism of intron excision takes place in the nucleus and is carried out by a multicomponent ribonucleoprotein complex – the splicosome.

1.2.1 Splicosome

The splicosome consists of five splicosomal small nuclear RNAs (snRNA) (U1, U2, U4/U6 and U5) and a large number of proteins including so called 'Sm-' core proteins, serine-arginine dipeptid family (SR-proteins) and further associated factors. Additional proteins are also found within this complex but their functions are not clear. The splicosome recognizes the splicing donor (exon n/ 5'-intron) and the splicing acceptor (intron-3'/ exon n+1). Within the splicosome two *trans* esterifications take place to produce the mature mRNA and excise the intronic region as lariat (Cartegni et al., 2002; Graveley, 2000).

1.2.2 Mechanism of splicing

Splicing consists of 4 main steps termed E, A, B and C. The nuclear pre-mRNA introns have in common only GU…AG/GURAGU nucleotides at 5' and 3' ends.
In the E-complex (E for early) U1 snRNA is bound to the 5' splice site, U2AF (AF=auxiliary factor) to the poly-pyrimidine tract and 3'AG and SF1/mBBP to the branchpoint (YNYURAC) sequence of the intron. The complex contains additional members of the SR protein family which recognize and bind on pre-mRNA sequences called exonic or intronic splicing enhancer (ESEs or ISEs) and silencer (ESSs or ISSs) (Black, 2003; Graveley, 2000; Maniatis and Tasic, 2002).
In the second step ATP is required to form the A complex in which U2 snRNP binds and defines the branchpoint (Michaud and Reed, 1991). The splicosome is assembled (B-complex) when U4/U6 and U5 snRNP joins the A complex and U5 snRNP recognized the 3'splice site (Chiara et al., 1997; Chua and Reed, 2001). Between complex E and B the intron ends are brought together to form a lariat (figure 1).

The mature splicosome (C-complex) now performes two rounds of catalysis to produce mRNA and the intronic lariat (Lewis et al., 1995; Michaud and Reed, 1991). Components that undertake the transesterification have not been directly identified. What is known is that first a free OH group (provided by an adenine nucleotide) attacks the exon-intron junction (of the donor) in a first transesterification reaction (Query *et al.*, 1996). An intermediate called lariat is formed (figure 1). The second splicing reaction follows rapidly. Binding of U5 snRNP to the 3'splice site is necessary for this reaction. The OH created at the 3' end of exon 1 attacks the intron-exon junction (of the acceptor). This results in the mature mRNA and the intronic lariat which gets degraded. The splicosome is then disassembled (Lewis et al., 1995; Sharp, 1994).

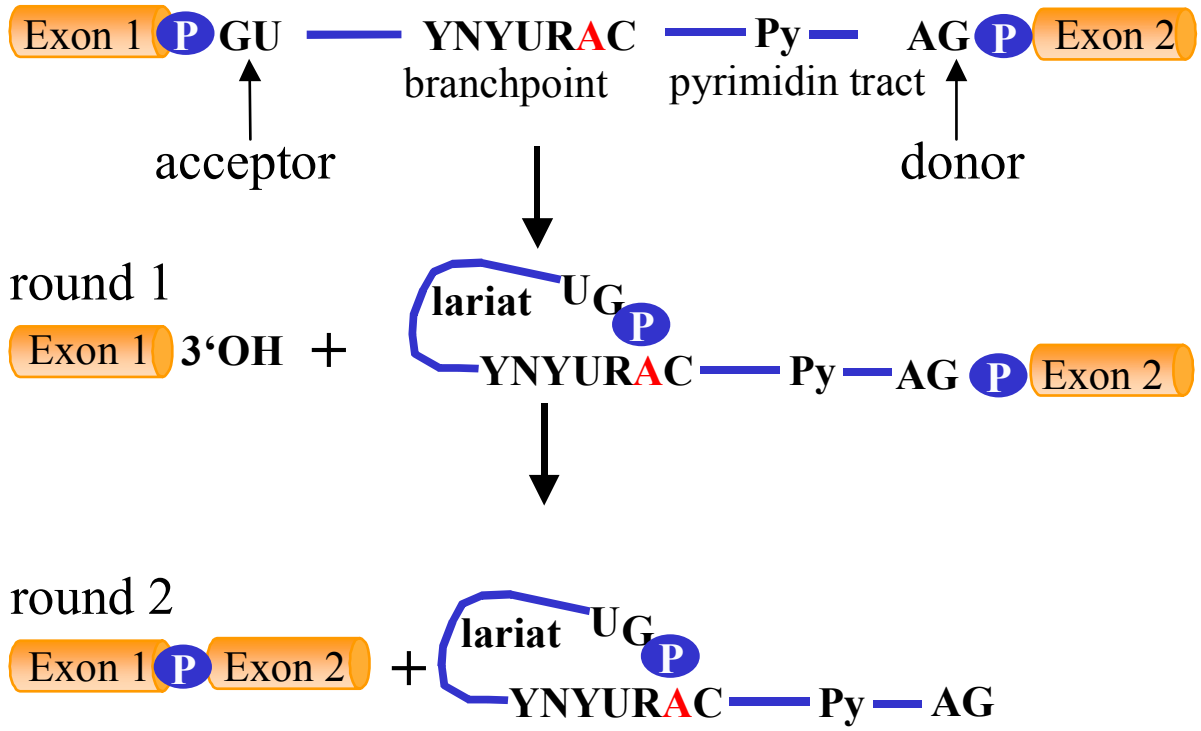


Figure 1: The mature splicosome (C-complex) undertakes two rounds of catalysis to produce mRNA and the intronic lariat.

1.2.3 Alternative splicing

Alternative splicing is a mechanism joining different 5' and 3' sites compared to the conservative splicing process. Thus multiple mRNA species are resulting encoding for proteins showing similar, converse or no function. A splicing error that adds or removes a single nucleotide will disrupt the open reading frame of an mRNA. Aberrant splicing generates variability that can insert or remove amino acids, shift the

reading frame, introduce termination codons, affect gene expression by removing or insertion of regulation elements for mRNA stability or protein localization (Faustino and Cooper, 2003).

It is estimated that about 23000 human genes exist, of which 40-70% show alternative transcripts (Brett et al., 2000; Johnson et al., 2003; Kan et al., 2005; Mironov et al., 1999; Modrek et al., 2001). Studies have identified about 30000 alternative spliced transcripts doubling the number of gene products (Lee and Roy, 2004) within humans. Organisms showing less complexity have reduced numbers of alternative splicing events with 40% of *Drosophila* genes showing one or more alternative exons, the worm *Caenorhabditis elegans* showing a lack of splicing and *S. cerevisae* has only several known regulated splicing events (Blencowe, 2006; Hodgkin, 2001; Kim et al., 2004; Stolc et al., 2004). Alternative splicing of pre-mRNA is one of the biological functions to increase diversity and is a regulatory component as different splicing transcripts can be seen in various tissues (Graveley, 2001; Johnson et al., 2003). Numerous binding proteins influence positively or negatively the use of splicing sites by binding to recognition sites in the exon or intron region (Black, 2003; Cartegni et al., 2002). These splicing signals can be targeted by mutations. It was found that around 15% of point mutations, which lead to genetic diseases cause alternative splicing defects (Krawczak *et al.*, 1992). The significance is supported by the annotation of around 54000 mutations in the Human Gene Mutation Database (http://www.hgmd.org/). Effects on splicing can be split up in *cis* and *trans* splicing. *Cis* elements affect splicing on one gene and *trans* elements affect splicing on multiple genes by disrupting components of the splicing machinery (Novoyatleva *et al.*, 2006).

The most common type of alternative splicing involves alternative cassette exons, which accounts for around one third of all aberrant splicing events (figure 2a). Second frequent are alternative 5' and 3' splice sites responsible for at least one quarter of all known aberrant splicing events (figure 2b). Alternative upstream or downstream exons are also quite frequent with 10 and 6% of all splicing events (figure 2c). Fourth most common are retained introns (figure 2d) (Blencowe, 2006; Van Gelder et al., 1990).

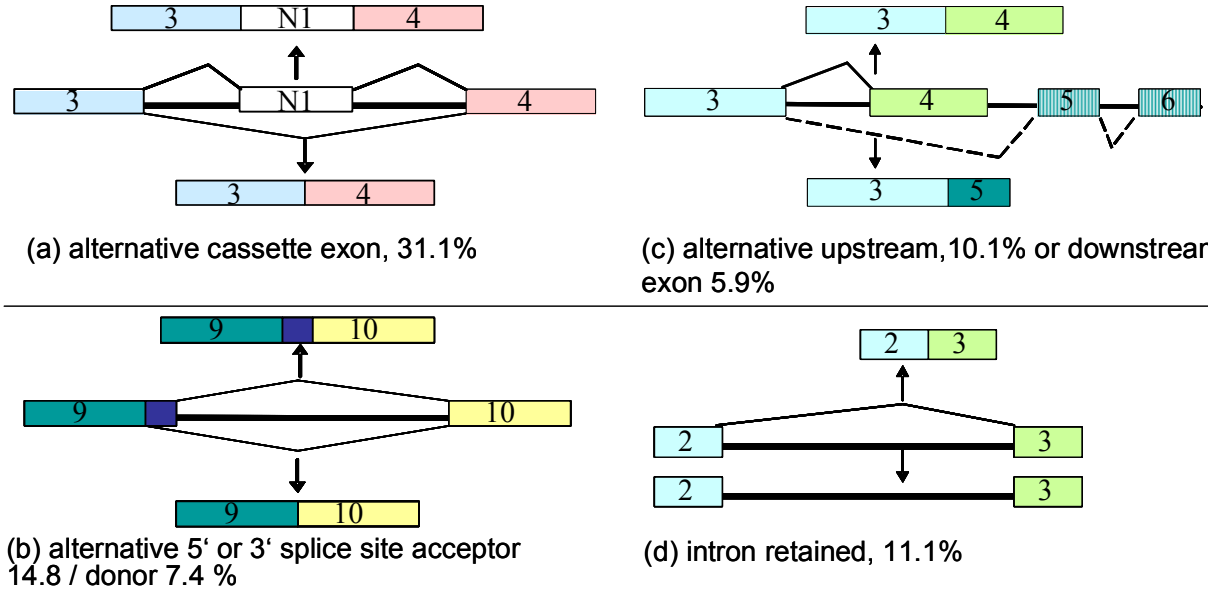


Figure 2: Representation of aberrant splicing events found within the human genome. Data taken and modified from a study performed by Watahiki et al. (2004).

1.2.4 Alternative splicing within pseudogenes

Pseudogenes are sequences that have close similarities to one or more paralogous functional genes and that presumably lack a functional product. Up to 15000 pseudogenes were found in the human genome in recent studies (Bischof et al., 2006; Khelifi et al., 2005; Yao et al., 2006; Zhang et al., 2006). Most pseudogenes are thought to be transcriptional silent but transcribed pseudogenes have also been identified (Balakirev and Ayala, 2003; Nelson et al., 2004; Zhang and Gerstein, 2004; Zhang et al., 2003). In a recent study it was found that roughly 14-16% of all pseudogenes from databases Vega and Ensemble have supporting expression evidence (Yao et al., 2006). As pseudogenes are usually not translated they are not subject to selection and hence reflect the pattern of naturally occurring mutations. These spontaneous mutations can lead to additional splicing variants in the pseudogenes not present in the gene. So far no comprehensive investigations were preformed comparing splicing patterns from genes and corresponding pseudogenes.

1.2.5 Human disease and adverse drug reaction (ADR) caused by alternative splicing

Alternative splicing can lead through previously described mechanisms to disease or adverse drug reactions. By disruption of a constitutive splice site the effect is often loss of function of the alternative spliced mRNA. In about 25-35% of all cases alternative splicing introduces premature termination codons (PTCs) (Novoyatleva et al., 2006). This leads in most cases to nonfunctional protein or loss of function due to nonsense-mediated decay (NMD). That can either lead to total loss of translation of this protein or to a change of the gene specific splicing patterns. Changing of splicing patterns of mRNA transcripts can have effects on tissue, developmental stages and increased or decreased enzyme activity of a translated enzyme.

For some members of the drug xenobiotic metabolizing Cytochrome P450 family alternative splicing is described and well studied. In drug metabolism, the most frequent null-alleles of *CYP2D6* as well as of *CYP2C19* harbor mutations at splice acceptor sites that prevent the expression of full-length functional proteins leading to the poor metabolizer phenotype in 5 to 10% of Europeans. Similarly, the polymorphism leading to almost complete lack of expression of *CYP3A5* in almost 90% of Caucasians has recently been elucidated to be mainly due to a splicing defect of the common **3* allele (Hustert et al., 2001; Kuehl et al., 2001; Toscano et al., 2006).

In a large number of human diseases alternative splicing plays a role as put together in the following examples: familial isolated growth hormone deficiency type II, frasier syndrome, frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17), atypical cystic fibrosis and genes associated with malignant disease (in genes such as *LKB1*, *KIT*, *CDH17*, *KLF6* and *BRCA1*) (Faustino and Cooper, 2003; Novoyatleva et al., 2006). Those examples can be explained by pre-mRNA mutations that affect usage of an alternative splice site. The variation in this *cis* element leads to a shift of the natural mRNA or protein to the unfunctional aberrant pre-mRNA or protein (Faustino and Cooper, 2003). Examples for diseases due to defective *trans* splicing are retinitis pigmentosa, spinal muscular dystrophy and certain types of neoplasmia and malignancy (in genes such as *Ron*, *RAC1* and *CD44*) (Faustino and Cooper, 2003; Novoyatleva et al., 2006; Srebrow and Kornblihtt, 2006). Interestingly, it was shown that some gene transfer methods like antisense nucleic acids (siRNA or short bifunctional oligonucleotides) and low molecular weight drugs (like

benzyladenine, polyphenol, sodium butyrate, valproic acid, sodium vanadate, indoprofen and indole derivatives) can interact with the RNA to yield increased levels of correctly spliced pre-mRNA in certain genes as recently reviewed (Novoyatleva et al., 2006).

1.2.6 Methods for detecting alternative splicing

There are three major ways of studying alternative splicing. The gold standard is based on cloned cDNA sequencing (Gerhard et al., 2004; Strausberg et al., 2002). Sequence based approaches and secondly computational analysis of alternative splicing using cDNA and expressed sequence taq (EST) data provide evidence of a vast number of alternative splicing isoforms. Limitations are the incompleteness and fragmentation of the EST and genomic sequence data (Modrek and Lee, 2002).
Secondly conventional analysis of RT-PCR can be performed with the limitation that many variants cannot be amplified, due to the placement or design of the primer pairs.
A third method to detect alternative splicing is based on microarrays. From 2001 it was shown that it is feasible to monitor alternative splicing events by placing specifically designed splicing probes on microarrays (Clark et al., 2002; Hu et al., 2001; Johnson et al., 2003; Shoemaker et al., 2001; Yeakley et al., 2002). Recently rather comprehensive genome wide arrays were published in a limited number of studies around the Affymetrix platform (Ule et al., 2005; Wang et al., 2003b), the Rosetta platform (Blanchette et al., 2005; Castle et al., 2003; Johnson et al., 2003; Pan et al., 2004) or the Exonhit platform (Fehlbaum *et al.*, 2005). These studies showed the capability to use microarray platforms with the focus on global detection and quantification of transcribed segments of any expressed gene. To finally assess the association of monitored alternative splicing transcripts, their formation via *cis* or *trans* acting elements and their consequences for phenotypes, extensive molecular studies are necessary.

1.3 Microarray platforms

DNA microarray technology has its origin in Edwin Southern's method described in his landmark article "Detection of specific sequences among DNA fragments separated by gel electrophoresis" published 1975 (Southern, 1975). In this publication the sequence-specific hybridization to probe genomic DNA is described. The next big step towards a microarray was hybridization analysis including "dot" blots and "slot" blots (Ristaldi et al., 1989; Wu et al., 1989). Southern again was the first to describe parallel *in situ* immobilisation of oligonucleotides as means of generating an oligonucleotide array on a solid support for highly parallel hybridization analysis (Southern and Maskos, 1994). Microarrays are distinguished from reverse dot blots by their high probe density and miniature size on solid impermeable, non porous substrate.

DNA microarray technology since then was shown to revolutionize life science data acquisition in many fields including genomic mapping (Wang *et al.*, 1998), clone library screening, gene expression profiling (Lockhart *et al.*, 1996), large-scale single nucleotide polymorphism (SNP) analysis (Mockler and Ecker, 2005) and to unravel gene functions, pathogenetic mechanisms, metabolic routes and the effects of drugs on these, to refine diagnostic classification and prognostic indexes, and to find new targets for therapy (Grimm et al., 2004; Jain, 2000; Marshall, 2003).

1.3.1 Principle of array based methods

The known nucleic acid, which is applied to analyse the unknown target sequence, is called the probe. Probes need to be specific and all have to be selected and tested for the application under the same hybridization conditions. DNA probes are immobilized in a highly parallel addressable format (Lipshutz *et al.*, 1999). The support consists of either glass, silicon, or a plastic substrate. Usually the support is coated with reactive groups (most common are poly-L-lysine, aldehyde or epoxy groups) to bind the target DNA via reactive groups of the nucleotides or with additionally synthesised exposed groups (e.g.NH_2). The nucleotides nearest to the surface are less accessible for hybridization of the target. Therefore spacer molecules can be used to increase the hybridization yield (Shchepinov et al., 1997; Southern et al., 1999).

The method of generating microarrays involves the use of photochemistry and photolithography for *in situ* oligonucleotide synthesis (Pease *et al.*, 1994) or piezoelectric nozzles and inkjet heads using surface chemistry directly on the substrate (Blanchard A.P *et al.*, 1996). Another method uses a version of non-contact/ contact printing (a way of micropipetting) mRNA, cDNA or oligonucleotides on surface (non- or covalently bound) (Cheung et al., 1999; Heller, 2002; Schena et al., 1995). This so called micropipetting can be performed with pins using a capillary slot, a ring and ping system, or inkjet dispensing systems.

For improved sensitivity target DNA or cRNA is amplified, labelled and afterwards immobilized on the microarray. The target DNA can be either DNA synthesised by polymerase chain reaction or RNA synthesised by one or more rounds of *in vitro* transcription. In addition T7-promoter can be included during cDNA synthesis to transcribe cRNA linearily with RNA polymerase (Van Gelder et al., 1990). In most cases labelling is performed with fluorescence dyes but also radioactive as well as silver or gold precipitations were described (Alexandre et al., 2001; Southern, 2001).

For maximal discrimination and perfect matched probe/ target duplex the hybridization reaction has to reach an equilibrium state which is dependent on the ionic strength and pH of the hybridization buffer, Tm of the duplexes, the hybridization temperature and time as well as the stringency of the following washing steps. All steps of microarray handling are put together in figure 3.

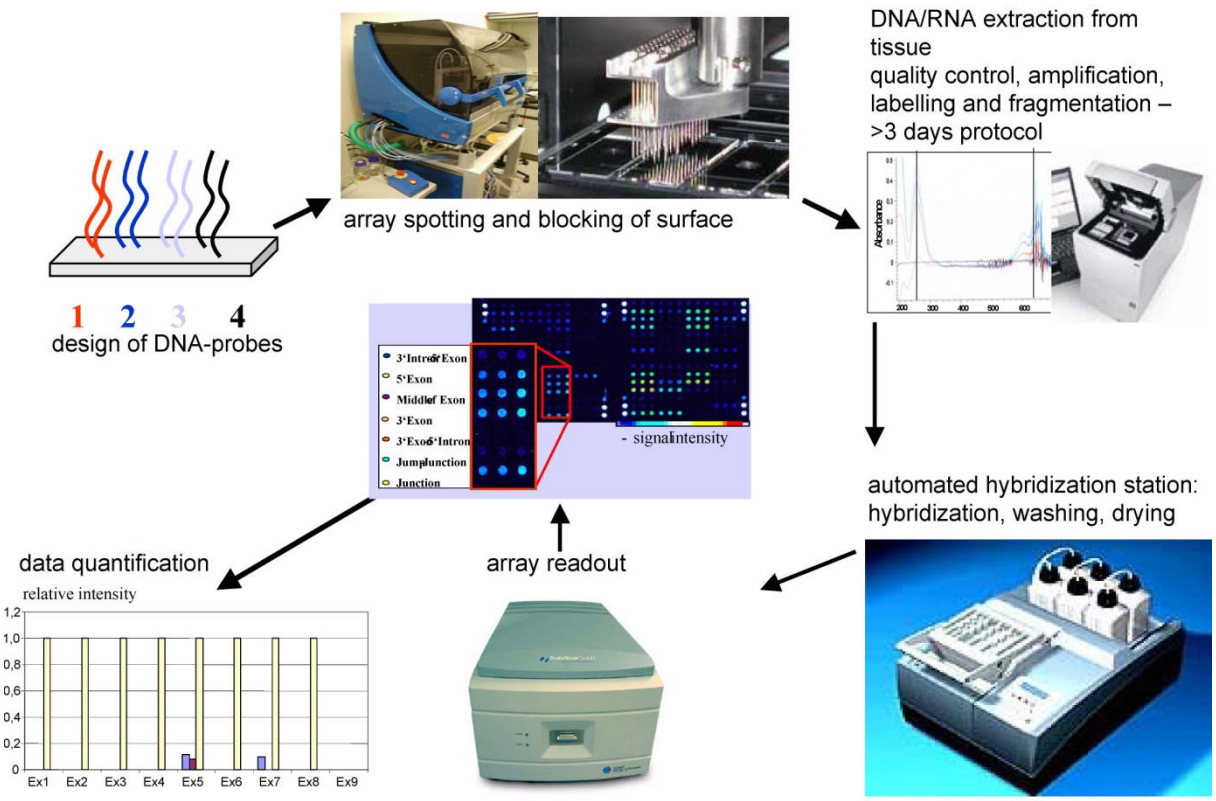


Figure 3: Steps of microarray experiments, which need to be performed and optimised.

1.3.2 Splicing microarray platform

In contrast to expression microarrays like the well-known Affymetrix GeneChip for detection of alternative splicing not just one but different expression values for each exon and intron need to be distinguished (Li and Wong, 2001).

Especially outliers who are ignored by conventional software might indicate alternative splicing events. In case of alternative splicing the presence and connection of alternative exons needs to be elucidated by specific probe design and special designed probes have to be used in combination with statistically analysis as illustrated in figure 4 (Cuperlovic-Culf et al., 2006; Le et al., 2004).

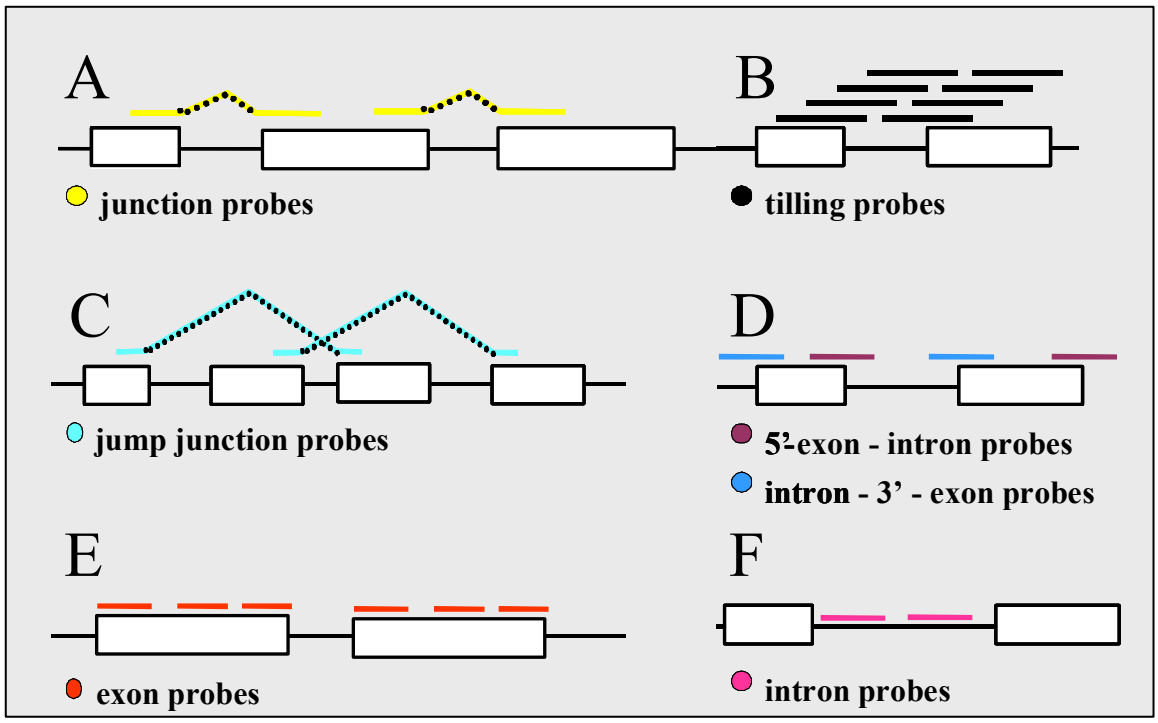


Figure 4: Special probes designed for alternative splicing platforms
A) depicts the placement of junction-probes (exon n to exon n+1) , B) tilling probes, C) jump junction probes (exon to exon n+2), D) intron-3'-exon and 5'exon-intron probes, E) exon probes and F) intron probes.

Probes placed on the exon-intron boundaries or junction probes (complementary to exon-exon junctions) were shown to be capable to detect exon boundaries and exon-exon junctions (Castle et al., 2003; Johnson et al., 2003). Tilling arrays were shown to predict exon structures with probes in the length of 50-60 nucleotides.
Alternative splicing arrays are so far dominated by two main commercial technology platforms – the Affymetrix gene technology array based on 25 nt. exon and intron control probes and the Agilent/ ExonHit platform based on 24 nt. exon and junction probes (Fehlbaum et al., 2005). Both are impressive platforms including probes for all genes of the human genome or in the case of Agilent/ ExonHit for a huge number of genes.
The results of those alternative splicing arrays are promising and it showed that this method is excellent for studying alternative splicing (Kan et al., 2005; Kirschbaum-Slager et al., 2004; Kirschbaum-Slager et al., 2005; Relogio et al., 2005).

1.3.3 Splicing microarrays and the existence of pseudogenes

A major problem for the interpretation of data from splicing microarrays is the existence of pseudogenes. The previous splicing microarray studies paid no attention to the existence of pseudogenes a phenomenon of special relevance for drug metabolizing enzymes and transporters. In humans the *CYP* gene superfamily consists of 57 functional genes and 58 highly homologous pseudogenes (Nelson et al., 2004). For some of the pseudogenes including *CYP2D7* (see below) expression in the liver was shown on the mRNA level (Endrizzi et al., 2002; Gaedigk et al., 2005). This emphasizes the necessity for specially designed splicing arrays in order to monitor the splicing patterns of the functional gene rather than a mixture of gene and pseudogene mRNA transcripts.

1.4 Interindividual variability in drug metabolism and response

Polymorphisms in metabolizing enzymes and transporters can be valuable biomarkers for variable pharmacokinetics, pharmacodynamics, ADR and clinical efficacy of numerous drugs. The systematic investigation of genetic factors on drug response is termed Pharmacogenetics.

Highly variable expression and function of these isozymes, both inter- and intraindividually, constitutes a major determinant for unpredictable drug and drug-metabolite plasma concentrations and can lead to unforeseen drug responses including over-reaction, toxicity, or lack of response (Meyer, 2000). Genetic polymorphisms in cytochrome P450 and other drug metabolism enzymes and transporter genes are an important source of variation, but induction/ repression or inhibition by drugs and other xeno- or endobiotics, biological or physiological conditions including sex, age, disease, and many other confounding factors often limit or modulate their penetrance. There are hundreds of single nucleotide polymorphisms (SNPs) and mutations within the drug metabolizing P450 genes and several of them were shown to induce alternative splicing.

During the last 30 years of pharmacogenetic research many genetic variations were found to explain inter-individual differences in drug response – with the earliest examples of human Cytochrome CYP2D6 (Eichelbaum et al., 1975; Eichelbaum et al., 1979) and drug-conjugating enzymes such as N-acetyltransferase, thiopurine methyltransferase and glutathione S-transferase (Koch, 2004; Weinshilboum and Wang, 2004).

The understanding of interindividual differences and the knowledge of the genetic background of patients gives the opportunity to individualized therapy with respect to the choice of drug and dose (Weinshilboum and Wang, 2004; Ingelman-Sundberg and Rodriguez-Antona, 2005).

1.4.1 The cytochrome P450 enzyme family

In humans 27 cytochrome P450 subfamilies with a total of 57 functional genes and 58 pseudogenes were identified (http://www.cypalleles.ki.se/) (Nelson et al., 2004).

About one dozen P450 isozymes of families CYP1, CYP2 and CYP3 are collectively responsible for most phase I biotransformations of drugs and other xenobiotics in

human liver. These P450 enzymes catalyze over 90% of all drug and xenobiotic metabolism pathways in humans (Ingelman-Sundberg, 2004; Lewis, 2004). All other subfamilies mainly metabolize endogenous substances like fatty acids, steroid hormones, prostaglandins as well as vitamin A and D.

The name Cytochrome P450 (CYP) originates from the early observations describing these enzymes as pigments (P) with a characteristic blue "Soretpeak", i.e. an absorbance near 450 nm in the presence of a reducing agent (with sodium dithionite) and carbon monoxide (Klingenberg, 1958). Individual CYP enzymes are designated by an arabic number for the gene family, a capital letter indicating the subfamily and an additional number for the gene (http://www.cypalleles.ki.se/). P450 proteins within prokaryotes are found in the cytoplasm. Within eukaryotes they are embedded in the membrane of endoplasmatic reticulum. The proteins consist of between 450 to 560 amino acids. The iron of the prosthetic heme group is bound to the P450 protein via a thiolate ligand derived from a highly conserved cysteine residue.

The typical reactions performed by cytochromes P450 are hydroxylations, N-, O-, S-dealkylations, N-, S-oxidations, deaminations, dehalogenations, desulfatations, epoxidations and peroxidations. The principal reaction referred to as monooxygenation as only one of two oxygens is incorporated into the substrate. The cytochrome P450 is the last part of an electron transfer chain. Via several steps one H^+ molecule is removed from the substrate, an O-molecule is covalently bound on the substrate and this way the product ROH is built, which is released by the cytochrome enzyme (Coon et al., 1992; Hlavica, 2004; Yasui et al., 2005).

1.4.2 Cytochrome CYP2D6

CYP2D6 is a highly polymorphic enzyme that metabolizes nearly 20% of all clinical used drugs (Zanger *et al.*, 2004). Four distinct drug oxidating phenotypes have been defined on the basis of activity: poor metabolizer (PMs), intermediate metabolizer (IMs), extensive metabolizer (EMs), and ultrarapid metabolizer (UMs).

CYP2D6 was the first Cytochrome P450 enzyme in which a genetically enzyme defect was discovered in the 1970s. Two research groups in England and in Germany observed that some volunteers participating in pharmacokinetic studies with debrisoquine (Mahgoub *et al.*, 1997) and sparteine (Eichelbaum et al., 1975), respectively and developed unexpected adverse reactions. In both cases it was shown that these individuals had a substantially impaired capacity to oxidize the drug

and that this metabolic deficiency is inherited as an autosomal recessive trait. The gene for *CYP2D6* is localized together with two highly homologous pseudogenes *CYP2D7* and *CYP2D8P* on chromosome 22 (Gonzalez et al., 1988; Kimura et al., 1989). More than 60 distinct alleles which result in CYP2D6 protein with normal, increased, decreased or complete lack of function have been documented on the CYPallele nomenclature homepage (http://www.cypalleles.ki.se) (Sim and Ingelman-Sundberg, 2006). The individually inherited allele combination (genotype) determines the drug oxidation phenotype, which can be assessed *in vivo* by measuring the metabolic ratio (MR), i.e. the ratio between the amount of unchanged drug and drug metabolite of a suitable probe drug (e.g. sparteine, debrisoquine, dextromethorphan) excreted in the urine within a certain time. Most of the CYP2D6 metabolised drugs contain a basic nitrogen, including antiarrhythmics (e.g. propafenone), antidepressants (e.g. amitriptyline, venlafaxine), antipsychotics (e.g. thioridazine), beta-blockers (e.g. metoprolol), opioids (e.g. codeine) and many more (Zanger et al., 2004). Predictive genotyping was shown in a recent publication to be beneficial for treatment in 30-40% of all CYP2D6 substrates, that is for about 7-10% of all drugs clinically used (Ingelman-Sundberg, 2005). So for optimal drug treatment correctly assignment of alleles and understanding of their phenotype is necessary.

1.4.2.1 Cytochrome *CYP2D6* alleles

Between 5 and 10% of Caucasians carry two non-functional null-alleles, of which about 20 are known to date. Most null alleles with SNPs lead to interruption of the reading frame (*CYP2D6 *3*, **6*, **8*, **13*, **15*, **16*, **19*, **20*, **21*, **38*, **42* and **56*) or interfere with correct splicing (*CYP2D6 *4*, **11*, **41* and **44*). Only a few non-functional alleles encode a full mRNA transcript and lead to a non-functional protein (*CYP2D6 *7*, **12*, **14* and **18*). Furthermore larger deletions or interruptions of the open gene frame exist in 2 alleles (*CYP2D6*5* and **16*) (Bradford, 2002; Ingelman-Sundberg, 2005; Zanger et al., 2004) (see the http://www.cypalleles.ki.se). The most frequent null-allele of *CYP2D6* is **4*, which occurs at a frequency of 20 to 25% in Caucasians (Kagimoto *et al.*, 1990).

Individuals with at least one allele encoding a functional protein are either EM, IM or UM. The quantitative differences among these individuals are extremely large. Most individuals carry either one or two functionally normal alleles (in most cases either *CYP2D6*1* or **2*, or rare ones like **33* or **35*) (Zanger et al., 2004).

Higher than normal activity UMs result from inheritance of one or two alleles with multiple functional copies, which occur in European populations with allele frequencies of around 1 to 10% (Johansson *et al.*, 1993) and with higher frequencies in certain Arabian and Eastern African populations (Ingelman-Sundberg, 1999). A phenotypically distinct subgroup of about 10 to 15% of Caucasians with metabolic ratio between those of EMs and PMs are commonly termed IMs (Zanger et al., 2004). The most common intermediate alleles are *CYP2D6*9* and *CYP2D6*41* with allele frequencies of 1-2% and 8% respectively, among Caucasiens.
In other ethnicities, the frequencies of variant alleles and of the resulting phenotypes are different. Thus, in Asians and Africans null-alleles are more rare than among Caucasians whereas alleles with reduced function, i.e. **10* in Asians (Mizutani, 2003) and **17* in Africans (Gaedigk *et al.*, 2002), are more common.

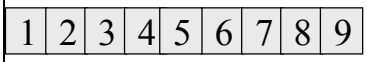
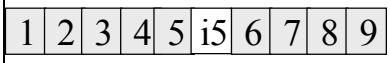
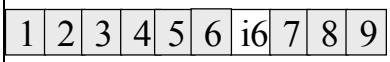
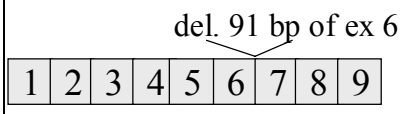
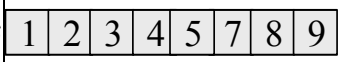
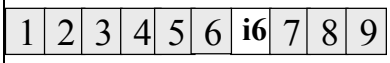
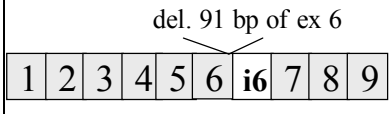
1.4.2.2 Splicing in *CYP2D6*

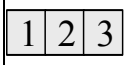
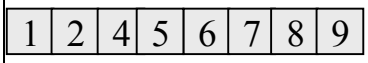
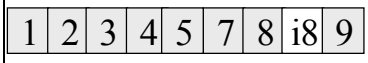
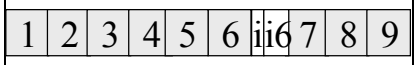
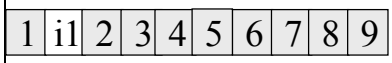
Within the literature 13 different splicing variants were described within *CYP2D6* and *CYP2D7* (table 1), which occur at various amounts in human liver and some in breast and brain tissue. Both the gene as well as the pseudogene is alternatively spliced, what complicates the study of aberrant splicing in both genes.
The phenotypes of two very common alleles, *CYP2D6*4* which is present in 20-25% and *CYP2D6*41*, in around 8% of Caucasians can be explained by splicing mechanisms. The key mutation 1846G>A in allele **4* changes the splice acceptor site of intron 3 such that it results in a shift of reading frame and ultimately a premature stop codon. Inheritance of two **4* alleles (or any other combination of two null alleles) leads to the CYP2D6 PM phenotype (Kagimoto et al., 1990).
The allele *CYP2D6*41* carries the key mutation 2988G>A located at nucleotide +39 of intron 6 (Raimundo et al., 2004; Toscano et al., 2006). 2988G>A leads to decreased amount of the normal transcript and substantially increased amount of a splice variant lacking the entire exon 6. This variant transcript harbours a stop codon at position 291 of the mature mRNA and can thus not contribute to functional CYP2D6 protein. Thus, the 2988G>A shifts the balance between the functional and a non-functional splice product towards the non-functional variant (Toscano et al., 2006). These two examples, which represent the two most common alleles with decreased function among Europeans, emphasize the relevance of alternative

splicing mechanisms for polymorphic drug metabolism (splicing in *CYP2D6* summarized in table 1).

Table 1: Compilation of transcript variants of *CYP2D6*

Shaded boxes represent exonic and white boxes intronic sequences. If possible nomenclature of splicing transcripts was used as described before. Numbers origin from sequence: AY545216 with ATG start codon set as 1. NP = normal product, SV = splicing variant.

name/ allele	mRNA transcript	description	tissue/ gene	literature
NP		exon 1-9: mRNA transcript that corresponds to the reference cDNA sequence		CYP2D6: AY545216 starting at ATP CYP2D7: M33387
SVa		inclusion of intron 5	liver/ brain *CYP2D6* and *CYP2D7*	Gonzales et. al. 1988 Gaedigk et al. 2005 Denson et al 2005
SVb		inclusion of intron 6 (in some cases in combination with intron 1 and 5 inclusion)	liver/ breast/ brain *CYP2D7*	Gonzales et al. 1988 Woo et al. 1999 Gaedigk et al. 2005 Denson et. al 2005
SVb'		lack of 91 bp in exon 6: 2860-2950 (*CYP2D6)*	liver/ breast *CYP2D6* and *CYP2D7*	Gonzales et al. 1988 Woo et al. 1999
SVc (*CYP2D6 *41)*		lack of exon 6 cause of 2988G>A	liver/ breast/ brain *CYP2D6 and CYP2D7*	Huang et al. 1997 Woo et al. 1999 Raimundo et al. 2004 Gaedigk et al. 2005 Toscano et al. 2006
SVd		inclusion of 57 bp of intron 6:14381-14437 (*CYP2D7)*	breast *CYP2D7*	Huang et al 1997
SVe		inclusion of 57 bp of intron 6: 14381-14437, (*CYP2D7)* and 91bp deletion of exon 6: 14153-14243 (*CYP2D7)*	breast *CYP2D7*	Huang et al 1997 Woo et al 1999
CYP2D6 11*		point mutation leads to alternative splicing cause of 883G>C	various tissue	Kagimoto et al. 1990 Maerz et al. 1997

name/ allele	mRNA transcript	description	tissue/ gene	literature
*CYP2D6 *4B-*4L*		splicing defect – splicing degeneration after exon 3	*CYP2D6*	Kagimoto et al. 1990 Maerz et al. 1997
SV1		lack of exon 3 is predicted in *CYP2D6*14* (1758G>A) and *CYP2D6*8* (1758G>T)	liver *CYP2D6*	Rogan et al. 2003
SV2		lack of exon 6 inclusion of intron 8 (sometimes in combination with intron 1 inclusion)	brain *CYP2D6*	Gaedigk et al. 2005
SV3		alternative acceptor of exon 6 and alternative donor of exon 7	liver and brain *CYP2D7*	Gaedigk et al. 2005
SV4		inclusion of intron 1	liver and brain *CYP2D6* and *CYP2D7*	Gaedigk et al. 2005

1.4.3 Cytochrome CYP2B6

The human cytochrome CYP2B6 is also a highly polymorphic enzyme, which plays a major role in the biotransformation of an increasing number of endogenous and exogenous substrates (Hodgson and Rose, 2007; Turpeinen et al., 2006; Zanger et al., 2007). CYP2B6 is the only functional isozyme of the 2B subfamily in humans and the gene is located, together with the expressed pseudogene *CYP2B7*, within a 350 kb *CYP2ABFGST* gene cluster on chromosome 19 that contains genes and pseudogenes of the *CYP2A*, *2B*, 2F, *2G*, *2S* and *2T* subfamilies (Hoffman *et al.*, 2001). *CYP2B6* contributes about 5% to the total liver microsomal P450 content. Clinically used drug substrates include cytostatics (cyclophosphamide (Chang et al., 1993; Roy et al., 1999)), HIV drugs (efavirenz (Ward *et al.*, 2003), nevirapine (Erickson *et al.*, 1999)), antidepressants (bupropion (Kirchheiner *et al.*, 2003)), antimalarias (artemisinin (Svensson and Ashton, 1999)), anaesthetics (propofol (Court *et al.*, 2001)), and synthetic opioids (methadone (Kharasch *et al.*, 2004)) and several others (Turpeinen et al., 2006). Variability of *CYP2B6* expression in liver is high and protein levels can be 100-fold different between individuals (Gervot *et al.*, 1999). Like the rodent phenobarbital-inducible *CYP2B* genes, human *CYP2B6* is strongly inducible by numerous drugs and chemicals including rifampicin, barbiturates, cyclophosphamide, artemisinin, carbamazepine, metamizole, and efavirenz (EFV) as well as nevirapine (Faucette et al., 2007; Saussele et al., 2007; Wang and Negishi, 2003; Zanger et al., 2007). Potent inhibitors of CYP2B6 are the platelet aggregation inhibitor clopidogrel (Richter *et al.*, 2004) and the cytostatic triethylenethiophosphoramide (thioTEPA) (Turpeinen et al., 2006; Rae et al., 2002).
Compared to the model polymorphism of *CYP2D6*, which was discovered by clinical observations, *CYP2B6* polymorphisms were discovered following reverse genetics approaches with initial polymorphism screening and subsequent functional and clinical studies (Lang *et al.*, 2001). Over 100 DNA variations including numerous nonsynonymous mutations as well as silent, promoter and intronic changes were found in the *CYP2B6* gene, many of them showing extensive linkage disequilibrium giving rise to distinct haplotypes (Ariyoshi et al., 2001; Blievernicht et al., 2007; Hesse et al., 2004; Klein et al., 2005; Lamba et al., 2003; Lang et al., 2004; Xie et al., 2003; Zukunft et al., 2005). Taking the number of variants and their frequency in different populations into account, *CYP2B6* belongs to the most polymorphic human

cytochrome P450 genes. Additionally the influence of numerous confounding factors like sex, induction by drugs and other xenobiotics, as well as possible substrate dependent effects contribute to *CYP2B6* variability (Zanger et al., 2007).

As mentioned before CYP2B6 is the main enzyme (with little contribution by CYP3A4) involved in sequential oxidation of efavirenz resulting in 8-hydroxy- and 8-, 14-dihydroxyefavirenz. The non-nucleotide reverse transcriptase inhibitor efavirenz, which is often used in combination with nucleoside inhibitors in first-line treatment to reduce mortality in HIV patients, shows a narrow therapeutic window with concentrations above 4 µg are associated with CNS side effects and below approximately 1µg/ ml with increased risk of virological failure (Marzolini et al., 2001; Telenti and Zanger, 2007). The association of *CYP2B6* polymorphic alleles and EFV plasma levels was analyzed in different clinical trials and it was a gene dose effect found with higher plasma levels in carriers within the partial deficient allele **6* (Owen et al., 2006; Rodriguez-Novoa et al., 2005; Rotger et al., 2006). Based on this a dose adjustment in patients carrying this genotype was suggested (Haas *et al.*, 2005).

1.4.3.1 Cytochrome *CYP2B6* alleles

To date, 53 distinct haplotypes and 28 defined alleles have been described; most of them carrying one or more amino acid changes but numerous promoter and intron changes contribute to further variants and highly complex haplotype structures. Additionally many further SNPs are known with unclear haplotype (CYPallele nomenclature homepage http://www.cypalleles.ki.se/cyp2b6.htm). The alleles within *CYP2B6* can't be classified as easily as in *CYP2D6* in defined groups. The most common alleles within *CYP2B6* are **1* and **6* followed by **2*, **4*, **5*, **9*, **17* and the loss of function allele **18*, which was so far only found in African populations (Zanger et al., 2007). 30-45 percent of all individuals are carriers of **1* (Zanger et al., 2007) but unknown mutations maybe still included within allele **1*.

The most common variant allele is *CYP2B6*6* and occurs with high frequency across different populations, ranging from 14% in Koreans to over 40% in West Africans and Chinese and 62% in Papua New Guineans (reviewed in Zanger et al., 2007). The *CYP2B6*6* allele harbours two nonsynonymous mutations which also occur individually (alleles **4* and **9*). The first key mutation is located in exon 4, c.516G>T [Q172H] in combination with the mutation c.785A>G [K262R] in exon 5. This common allele recently obtained clinical relevance when CYP2B6 was described as

the major catalyst for oxidative metabolism of efavirenz (EFV) (Desta et al., 2007; Ward et al., 2003). This lead to subsequent pharmacogenetic studies in HIV patients, which consistently revealed strong associations between severely elevated EFV plasma concentration and homozygosity for the common *CYP2B6*6* allele [Q172H and K262R] and several minor loss of function alleles in various ethnic populations (Haas, 2006; Owen et al., 2006; Telenti and Zanger, 2007). Moreover, recently *CYP2B6* genotyping in Japanese HIV-infected individuals was successfully applied to reduce the therapeutic dose of EFV, which resulted not only in the expected pharmacokinetic changes but also in improvement of CNS-related side effects (Gatanaga *et al.*, 2007). Despite its established clinical genotype-phenotype relationship, the molecular basis of the **6* allele has remained unclear and various studies indicated possible effects at the level of transcription, splicing, protein stability, as well as substrate specificity and turnover. In a recent extended study it was shown that heterozygous and homozygous carriers of **6* allele indeed express up to four-fold less protein with concomitant changes in activity as compared to homozygotes of the reference allele *CYP2B6*1* (Desta et al., 2007). The reason for lower expression in liver remained however unknown.

1.4.3.2 Splicing in *CYP2B6*

A number of variant transcripts were described within the *CYP2B6* gene. First transcript variants were described by Miles et al (1989). A recent study observed additional variants that were suggested to be associated with certain polymorphisms (Lamba et al., 2003). However, all these studies were descriptive and they cover only parts of the gene. Most importantly the contribution of these variants to the ultimate hepatic activity phenotypes remained unclear.

1.5 Objectives

The primary objectives of this thesis is the molecular analysis of transcript diversity occurring by alternative splicing within the polymorphic cytochromes *2D6* and *2B6* and to find a new explanation of phenotypes with the help of alternative splicing.

The specific aims with respect to *CYP2D6* are the development of a splicing microarray to detect aberrant splicing transcripts within human liver tissue and the relative quantitation of the most common splicing transcripts. The splicing array will be the first existing microarray designed to specifically discriminate between gene and pseudogene specific splicing transcripts. The microarray should be capable to study and compare gene and pseudogene specific splicing patterns. Results will be confimed by RT-PCR and sequencing and compared to literature data.

With respect to *CYP2B6* the specific aims are the transfer of the newly designed splicing microarray platform from *CYP2D6* to *CYP2B6*. Splicing patterns in human liver tissue detected by the *CYP2B6* splicing array should be confirmed by alternative methods. The molecular mechanism leading to the CYP2B6*6 phenotype, showing reduced protein expression and activity, is not explained so far. This should be investigated with the help of the developed splicing array, the usage of a cloned recombinant plasmid systems and the study in cell systems.

To achieve this following resources and tools were available at the institute:
A collection of well-characterized pheno- and genotyped human liver samples from patients was available. These samples were suitable to study alternative splicing with RT-PCR, sequencing, and the newly developed microarrays and to draw connections to certain SNP’s and alleles.

2 Material and methods

2.1 Material

2.1.1 Biological reagents

Supplier	**Product**
Agilent (Böblingen,Germany)	DNA 1000 LabChipKit, RNA 6000 Nano LabChipKit
Applied Biosystems (Ambion & Tropix) (Darmstadt, Germany)	MEGAscript T7 kit, RNase H, Taq Man reverse transcription reagents, TaqMan universal master mix, Galacto-Star™ ß-Galactosidase Reporter Gene Assay (Galactone, Emerald enhancer)
Biolog (Bremen, Germany)	ddNTPs
Bio-Rad (Munich, Germany)	Bio-Rad protein assay, BSA standard protein
Biozym (Oldendorf, Germany)	Chill Out 14 Lipid Wax (red) for PCR
GE Healthcare/ Amersham (Freiburg, Germany)	dNTP Set, fluoroLink Cy3-dCTP, Cy3 Post-labelling reactive dye pack, thermo sequenase fluorescent, fluorescent labelled primer sequencing kit with 7-deaza-dGTP, rainbow coloured protein molecular weight markers
Invitrogen (Karlsruhe, Germany)	TOPO TA cloning kit for sequencing, 5x second strand buffer, 100 bp DNA ladder, 1kb Plus DNA ladder, Superscript III
Metabion (Planegg-Martinsried, Germany)	5' or 3' C_6-amino-modified oligonucleotide probes for splicing array, oligonucleotide probes with 5' Cy3/Cy5-modification (controlled by MALDI-TOF), PCR primers (desalted)
MWG (Ebersberg, Germany)	IR 800 labelled oligonucleotide probes for sequencing
New England Biolabs (Frankfurt, Germany)	*Acc*65I, *Bbr*CI, *Bsm*BI, *Bcg*I, *Bsg*I, 100x BSA (bovine serum albumin), DNA polymerase I, *E.coli* ligase, *Eco*NI, *Mfe*I, *Not*I, *Pme*I, *Sbf*I, *Ssp*I, T4 DNA polymerase
Promega (Mannheim, Germany)	RQ1 RNase-free DNase kit
Qiagen (Hilden, Germany)	Taq DNA polymerase, PCR buffer (10x), HotStar Taq master mix kit, nuclease free water, RNase free DNase set, RNeasy mini kit, QIAquick

Supplier	Product
	PCR purification kit, QIAquick gel extraction kit, QIAprep spin miniprep kit, QIAshredder, Qiagen plasmid maxi kit
Roche Applied Science (Mannheim, Germany)	shrimp alkaline phosphatase (SAP), *Eco*RI, expand high fidelity PCR system, Fugene HD transfection reagent, *Nhe*I, *Sac*I, *Spe*I, T4 DNA ligase, complete mini (protease inhibitor cocktail)
Solis Biodyne (Tartu, Estland)	termipol polymerase
Stratagene (Amsterdam, Netherlands)	site directed mutagenesis kit

2.1.2 Chemical reagents

Supplier	Product
Becton Dickinson, Heidelberg	bacto agar, bacto yeast, bacto trypton
Bio-Rad (Munich, Germany)	acrylamide/Bis (30:0.8)
Bruker Daltonik (Bremen, Germany)	acidic 3-hydroxypicolin
Cambrex, (Rockland, USA)	long range gel solution
Fluka (Deisenhofen, Germany)	ammonium acetate, milk powder (low fat), NADPH, $NADP^+$
Invitrogen (Karlsruhe, Germany)	agarose
Merck (Darmstadt, Germany)	ammoniumpersulfate (APS), liChrosolv (HPLC water), HCl, KCl, NaCl, disodiumhydrogenphosphate, Tween 20
Peqlab (Erlangen, Germany)	nexterion Spot I, nexterion Spot III
Promega (Mannheim, Germany)	passive lysis buffer, urea
Roche Diagnostics (Mannheim, Germany)	glucose-6-Phosphate, glucose-6-phosphate-dehydrogenase
Roth (Karlsruhe, Germany)	acetonitrile, ampicillin, IPTG, kanamycinsulfat, low melt agarose, methanol, tris-base, X-Gal
Sigma-Aldrich (Deisenhofen, Germany)	ß-mercaptoethanol, acetic acid, bromphenol blue, bovine serum albumin (BSA), dimethylsulfoxide (DMSO), ethylendiamintetraacetic acid (EDTA), ethanol (100% or 95%), ethidium bromide, ethylene glycol, glycerol, hydroxylamine, ponceau S-solution, sodiumdodecylsulfate (SDS), sodium acetate, sodium borate, sodium citrate·$2H_2O$, Na_2EDTA, NaOH, N,N,N-tetramethyl ethylendiamine, triton X-100

2.1.3 Equipment

Supplier	Designation	Equipment
ABgene (Hamburg, Germany)	microarray sealing PCR plates	gene frame seals 65µl thermo fast 384/96 PCR plates
Agilent (Böblingen, Germany)	RNA and DNA Analyser photometer HPLC-MS/MS spectrometry	Agilent 2100 Bioanalyzer UV-Visible Chem Station HPLC system (HP 1100)
Applied Biosystems (Darmstadt, Germany)	DNA-sequencer TaqMan	ABI Prism 7500 7500 Real Time PCR System
Becton Dickinson (Heidelberg, Germany)	tissue culture	tissue culture plate (12 well)
Berthold Technologies (Bad Wildbach, Germany)	luminometer	AutoLumat Plus
Biometra (Göttingen, Germany)	hybridization oven UV-crosslinker western blotting	OV5 hybridization oven Stratalinker UV Crosslinker 1800 fastblot B43
Bio-Rad (Munich, Germany)	electrophoresis of polyacrylamide gels	Trans-Blot Cell
Biorobotics (Cambridge, England)	microarray spotter	Microgrid II Microarrayer
Bruker Daltonik (Bremen, Germany)	Maldi-Tof-MS robotic workstation anchorChip ion trap mass spectrometer	Bruker Ultraflex MALDI-TOF masspectrometer Pure Disk MTP anchor chip 400/384 TF HCT plus
Eppendorf (Hamburg, Germany)	centrifuge thermomixer thermomixer slide adapter pipettes	5417C thermomixer comfort thermomixer slide adapter research, research pro
Greiner bio-one (Frickenhausen, Germany)	spotting plate (for microarray spotter)	384 well polystyrol (PS)-microplates
Heraeus (Hanau, Germany)	centrifuge	Biofuge 13 Biofuge 22 Biofuge pico
Hettich (Tuttlingen, Germany)	centrifuge	Hettich Rotana/S, 23 Hettich Universal 32 Hettich Universal 320R
Hirschmann Laborgeräte (Heilbronn, Germany)		glas capillaries
Infors-HT (Bottmingen, Switzerland)	incubator shaker for liquid bacterial culture	Aerotron
LI-COR (Lincoln, USA)	DNA-sequencer Infrared imaging system for western blot analysis	4300 DNA Analyser Odyssey

Supplier	Designation	Equipment
LTF Labortechnik (Wasserburg, Germany)	gel documentation system	UV-light-transilluminator
Memmert (Schwabach, Germany)	drying compartment	universal oven (Model U)
Millipore (Molsheim, France)	water purification systems	Milli Q Ultrapure
MJ Research Inc. (Waterdown, USA)	PCR cycler	PTC-200 thermal cycler PTC-225 thermal cycler
MP Biomedicals - Bio 101 Systems (Heidelberg Germany)	homogenisator cell lysis	FastPrep FP120 Lysing tubes with matrix D
NanoDrop Technologies (Wilmington, USA)	spectrophotometer	ND-1000 Spectrophotometer
Nunc (Wiesbaden, Germany)	cell flasks	cell flasks T-75 (75 cm^2)
Owl (Porthsmouth, USA)	agarose gel electrophoresis	easy cast model B2 (12x13 cm)
PeqLab (Erlangen, Germany)	agarose gel electrophoresis	perfect blue gel systems S (7x8 cm)
Peqlab (Erlangen, Germany)	microarray glass slides	nexterion slide E
PerkinElmer (Rodgau – Jügesheim, Germany)	array-scanner	ScanArray express microarray scanner
Sarstedt (Nümbrecht, Germany)	cell harvesting ß-galactosidase assay	cell scraper tubes 75x11.5 mm
Schärfe System (Reutlingen, Germany)	cell counter	Casy 1
Schleicher & Schuell (Dassel, Germany)	transfer membrane steril filtration	nitrocellulose sterilfilter (0.2 µm and 0.45 µm)
Schott (Mainz, Germany)	pH-meter	pH-meter CG840
Scientific Industries (New York, US)	vortexer	Vortex-Genie 2
Tecan (Crailsheim, Germany)	hybridization station	HS 400 hybridization station
Thermo Electron (Hanau, Germany)	centrifuge	Sorvall superspeed RC2B
Trentec Analysentechnik (Gerlingen, Germany)	HPLC column	Trentec-Reprosil-Pur 120-ODS3 column (50x23 mm I.D., 5 µm)

2.2 Buffers and solutions

2.2.1 Buffers for microarray experiments

Name	**Composition**	
20x SSC	sodium citrate*2 H_2O NaCl 1M HCl	0.3 M 3 M adjust pH to 7.0
Pseudo 20x SSPE – used for hybridization	NaCl sodium phosphate EDTA	3.6 M 0.2 M 20 mM
Rinsing solution 1	triton X 100 H_2O_{mp}	2 ml ad 2000 ml
Rinsing solution 2	HCl (37%) H_2O_{mp}	1 ml ad 2000 ml
Rinsing solution 3	KCl-solution 1mM H_2O_{mp}	200 ml 1800 ml
Blocking solution	ethylenglycol H_2O_{mp} HCl (37%)	500 ml 1500 ml 1 ml
Washing solution 1	20x SSC SDS (10%) H_2O_{mp}	200 ml 40 ml 1760 ml
Washing solution 2	20x SSC H_2O_{mp}	200 ml 1800 ml
Washing solution 3	20x SSC H_2O_{mp}	20 ml 1980 ml
2x spotting solution X	Na_2SO_4 $Na_2HPO*2H_2O$ H_2O_{mp}	44.1 g 44.6 g ad 1000 ml
2x spotting solution Y	spot I spot III (spot III= spot I +0.02% triton X)	1 part 3 parts (vol/ vol)
labelling buffer	dimethyloxide 1M $NaHCO_3$ (pH 9.0)	1:1 (vol/ vol)

2.2.2 Buffer for SDS page and western blotting

Name	**Composition**	
APS (10%)	ammoniumpersulfate H_2O_{mp}	1 g ad 10 ml
Laemmli sample buffer (5x)	Tris-HCl pH 6.8 glycerin SDS bromphenolblue ß-mercaptoethanol	60 mM 24% 2% 1% 14.4 mM
SDS 10%	SDS H_2O_{mp}	50 g ad 500 ml
Electrophoresis buffer (10x)	Tris-base glycine SDS H_2O_{mp}	150 g 720 g 50 g ad 5000 ml
Electrophoresis buffer (1x)	electrophoresis buffer (10 x) H_2O_{mp}	500 ml ad 5000 ml
Blotting buffer	Tris-base glycin SDS 10% H_2O_{mp} methanol	9 g 14.6 g 18.5 ml ad 4000 ml ad 5000 ml
TBS (10x)	Tris-base NaCl KCl H_2O_{mp} HCl	150 g 400 g 10 g ad 5000 ml adjust to pH 7.4
TBST	Tween 20 TBS (10x) H_2O_{mp}	10 ml 500 ml ad 5000 ml
Tris-HCl 0.5 M, pH 7.5	Tris-base H_2O_{mp} HCl	30 g ad 500 ml adjust pH to 7.5
Stacking gel (4%)	acrylamide/bis (30:0.8) Tris-HCl 0.5M (pH 6.8) SDS (10%) APS (10%) TEMED H_2O_{mp}	1.35 ml 2.5 ml 0.1 ml 0.1 ml 10 µl ad 6.1 ml
Resolving gel (10%)	acrylamide/bis (30:0.8) Tris-HCl 1.5M (pH 8.8) SDS (10%) APS (10%) TEMED H_2O_{mp}	10.0 ml 7.5 ml 0.3 ml 0.3 ml 30 µl 12.0 ml

2.2.3 Buffers and solutions for protein extraction

Name	**Composition**	
Complete-solution, EDTA-free	1 tablet complete mini H_2O_{mp}	 1.5 ml
Lysisbuffer for protein	Tris-HCl pH 7.5 M NaCl Triton X-100 (v/v) EDTA complete- solution (EDTA free (v/v))	50 mM 0.25 M 0.1% 5 mM 15%

2.2.4 Buffers and solutions for sequencing and agarose gel casting

Name	**Composition**	
Sequencing loading dye	formamide ethylenediamine tetraacetic acid (EDTA) new fuchsin red H_2O_{mp}	95% 0.5 mM 0.1 mg/ ml ad 100%
Sequencing gel	Urea Long ranger solution TBE (10x) DMSO H_2O_{mp} after filtration add APS (10%)	21 g 6 ml 5 ml 500 µl ad 50 ml 400 µl
DNA loading buffer (5x)	Ficoll (20%) bromphenolblue (0.5%) EDTA (0.5 M, pH 8)	874 µl 87.4 µl 38 µl
TBE (10x)	Tris boric acid EDTA H_2O_{mp}	162 g 27.5 g 9.3 g ad 1000 ml
TAE (50x)	Tris glacial acetic acid EDTA (0.5M, pH 8) H_2O_{mp}	242 g 57.1 ml 100 ml ad 1000 ml

2.2.5 Buffers and solutions for cytochrome-P450-activity measurements

Name	**Composition**	
0.1M sodiumphosphate pH 7.4	Na_2HPO_4 0.1 M NaH_2PO_4	0.1 M adjust pH to 7.4
0.1M Tris pH 7.5	Tris HCl	0.1 M adjust pH to 7.5

10x NADPH regenerating system (NADPH-RS)	$NADP^+$ Glucose-6-phosphate $MgCl_2$ Glucose-6-phosphate-dehydogenase dissolved in: sodiumphosphatebuffer pH7.4	5 mM 40 mM 50 mM 40 U/ ml
denatured microsomal protein	pooled microsomes denatured at 100°C for 5 min and stored at -20°C	with average protein concentration of 3.5 mg/ ml

2.3 Bacterial and eukaryotic cell culture

2.3.1 Bacterial cell culture

Supplier	***E. coli* strains**
Stratagene (La Jolla, USA)	XL-1 Blue
Invitrogen (Karlsruhe, Germany)	TOP 10

The following bacteria growth media was autoclaved after preparation:

Ampicillin solution	ampicillin ethanol (99%) H_2O_{mp}	50 mg/ ml 50% 50%
Kanamycin solution	kanamycin sulfate H_2O_{mp}	10 mg/ ml 100%
LB ("Luria-Bertani") medium	tryptone yeast extract NaCl H_2O_{mp}	10 g 5 g 5 g ad 1 l
LB + agar medium	LB medium agar	1 l 15 g
SOC	Trypton Yeast extract KCl H_2O_{mp} after autoclaving ad: MgCl (1 M sterilfiltrated) $MgSO_4$ (1 M sterilfiltrated) glucose(2 M sterilfiltrated)	20 g 5 g 0.18 g ad 970 ml 10 ml 10 ml 10 ml

2.3.2 Eukaryotic cell culture

Following eukaryotic cell lines were used:

COS-1: African green monkey kidney epithelial cell lines, kindly provided by Prof. Dr. Urs Meyer, Division Pharmacology/ Neurobiology Biozentrum, University of Basel, Switzerland

Huh7: human hepatoma cell lines, kindly provided by Epidauros AG, Bernried, Germany.

Following media were used for eukaryotic cell culture

Supplier	
Invitrogen (Karlsruhe, Germany)	DMEM (Dulbecco's modified eagle medium), MEM (minimum essential medium) L-glutamine penicillin-streptomycin liquid pyruvat
Biochrom (Berlin, Germany)	fetal bovine serum

The media for eukaryotic cells were supplemented with following ingredients before usage:

Media for COS-1 cells	DMEM supplemented with : fetal bovine serum L-glutamine penicillin streptomycin	 10% (v/v) 1% (v/v) 100 U/ ml 100 µg/ ml
Media for Huh-7	MEM supplemented with fetal bovine serum L-glutamine penicillin streptomycin	 10% (v/v) 1% (v/v) 100 U/ ml 100 µg/ ml

2.3.3 Expression vectors and plasmids

Following expression vectors were used or generated in this study:

Vector	**Properties**	**Origin/ Reference**
pCR4-TOPO	T7 and T3 RNA polymerase promoter, direct cloning of PCR products	Invitrogen, Karlsruhe, Germany
pCMV4	mammalian expression plasmid driven by CMV promoter	U.M. Zanger, IKP Stuttgart, Germany
pCMVß	ß-galactosidase expression vector	BD Biosciences Clontech, Palo Alto, USA
pBS SK+2D6	pBluescript SK+ vector containing the *CYP2D6* wild type cDNA	Evert et. al. 1997
pIKAT15	*CYP2B6*-cDNA vector	Lang et. al. 2004
pMHH-CYP2B6-TOPO-98	pCR4-vector containing splice variant SV1 of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-101	pCR4-vector containing splice variant SV8 of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-134	pCR4-vector containing	this study

Vector	Properties	Origin/ Reference
	conservative splice product of *CYP2B6* between exon 3-7	
pMHH-CYP2B6-TOPO-111	pCR4-vector containing splice variant SV7 of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-207	pCR4-vector containing splice variant SV9 of *CYP2B6* between exon 5-9	this study
pMHH-CYP2B6-TOPO-215	pCR4-vector containing splice variant λMP8 of *CYP2B6* between exon 5-9	this study
pMHH_CYP2B6_Mini*1	expression plasmid driven by CMV promoter containing the *CYP2B6 wild type* minigene	this study
pMHH_CYP2B6_Mini*4	expression plasmid driven by CMV promoter containing the *CYP2B6 *4* minigene	this study
pMHH_CYP2B6_Mini*3	expression plasmid driven by CMV promoter containing the *CYP2B6 *3* minigene	this study
pMHH_CYP2B6_Mini*9	expression plasmid driven by CMV promoter containing the *CYP2B6 *9* minigene	this study
pMHH_CYP2B6_Mini*6	expression plasmid driven by CMV promoter containing the *CYP2B6 *6* minigene	this study

2.3.4 Antibodies and recombinant proteins

Antibodies and recombinant proteins used in this study:

Supplier		
BD Biosciences (Frankfurt, Germany)	anti-human CYP2B6 (mouse)	2B6 monoclonal antibody
LI-COR (Lincoln, USA)	IR Dye 800 goat anti mouse IgG	2B6 secondary antibody
MoBiTec (Göttingen, Germany)	anti-ß-galactosidase *(E.coli)*	ß-gal monoclonal antibody
Sigma-Aldrich (Deisenhofen, Germany)	ß-galactosidase (*E.coli*) anti-ß-actin (human, mouse, rabbit)	ß-gal protein ß-actin monoclonal antibody
Gentest (Franfurt, Germany)	cytochrome P450 2B6 out of lymphoblastoid cells	CYP2B6 protein

2.3.5 Oligonucleotides used for PCR, sequencing and in vitro mutagenesis

Primer Designation	**5' to 3' Sequence**	**Primer Location**	**Gene Specificity**
Construction of *CYP2B6* minigenes			
2B6_3'ecoNItaq_E2_for 1	cctttctgaggttccgagagaaat at	Exon 2	*CYP2B6*
2B6_5'sacItaq_E4_rev1	gagctcaaacagctggccgaat acagagctgatgagtgaa	Exon 4	*CYP2B6*
CYP2B6-5F	gacagaaggatgagggaggaa	Intron 4	
2B6_spez_I8R_rev1	tcgttgtttttctcaagttggggata gt	Intron 8	*CYP2B6*
2B6_5'SbfI_Ex7_for1	cctgcagggccatacacagagg cagtcatctat	Exon 7	*CYP2B6*
***In vitro* Mutagenesis**			
2B6_Ex4_15631T_for	ccaccttcctcttccattccattacc gccaac	Exon 4	*CYP2B6*
2B6_Ex4_15631T_rev	gttggcggtaatggaatggaaga ggaaggtgg	Exon 4	*CYP2B6*
2B6_Ex5_777C_for	gaaaccctggaccccagcgccc ccaaggacctc	Exon 5	*CYP2B6*
2B6_Ex5_777C_rev	gaggtccttgggggcgctggggt ccagggtttc	Exon 5	*CYP2B6*
2B6_Ex5_18053G_for	ccccagcgcccccagggacctc atcgacac	Exon 5	*CYP2B6*
2B6_Ex5_18053G_rev	gtgtcgatgaggtccctgggggc gctgggg	Exon 5	*CYP2B6*
***CYB2B6* TaqMan Primer and Probe**			
TQ-CYP2B6for	gctgaacttgttctaccagacttttt c	Exon 4	*CYP2B6*

Primer Designation	5' to 3' Sequence	Primer Location	Gene Specificity
TQ-CYP2B6rev	gaaagtatttcaagaagccaga gaagag	Exon 5	*CYP2B6*
CYB2B6MGB	6-FAM-tgtattcggccagctgt-MGB	Exon 4 and 5	*CYP2B6*
RT-PCR primers			
cDNA2B6FE2	ttcttccggggatatggtgtga	Exon 2 and 3	*CYP2B6*
cDNA2B6RE7	tctccctgtagactctctctgca	Exon 7	*CYP2B6*
cDNACYP2B6FE5	ctgtttgagctcttctctggcttc	Exon 5	*CYP2B6*
cDNACYP2B6RE9	tcagcggggcaggaagcggatc tg	Exon 9	*CYP2B6*
T7-Primers			
CYP2D6-T7	aaacgacggccagtgaattgta atacgactcactataggcgctca ccaggaaagcaaagaca	Exon 9 including T7 promoter	*CYP2D6*
CYP2D7-T7	aaacgacggccagtgaattgta atacgactcactataggcgctca ccagaaagctgacgaca	Exon 9 including T7 promoter	*CYP2D7*
CYP2B6-T7	aaacgacggccagtgaattgta atacgactcactataggcgcgtgt cagatcgatgtcttctggg	Exon 9 including T7 promoter	*CYP2B6*
Sequencing Primer			
M13 forward	gtaaaacgacggccag		
M13 reverse	caggaaacagctatgac		
seqCYP2B6-RE2	atatactcccttccctgatgca	Intron 3	*CYP2B6*
seqCYP2B6-FE3	actcagagccttcttccaact	Intron 2	*CYP2B6*
seqCYP2B6-FE4	taggtgacagcctgatgttc	Intron 3	*CYP2B6*
seqCYP2B6-RE5	tcttctcacctctccatctt	Intron 5	*CYP2B6*
seqCYP2B6-FE6	tatacacagcaaggctacag	Intron 5	*CYP2B6*
seqCYP2B6-RE6	atttctgcagctcagaagga	Intron 6	*CYP2B6*
seqCYP2B6-FE7	gattacaggcatgagccaccat	Intron 6	*CYP2B6*
seqCYP2B6-RE7	attaagagaatccaggatgcc	Intron 7	*CYP2B6*
seqCYP2B6-FE8	ttttgtggagtgtgtggagggt	Intron 7	*CYP2B6*
PCR Primer			
2D6_Ex1_Maldi	acgttggatgatggggctagaag cactggtg	Exon 1	*CYP2D6*
2D6_rev3_Maldi	acgttggatgtcaccaggaaagc aaagaca	Exon 9	*CYP2D6*
5'exon3_rev	gcgcgtggccatgaagg	Exon 3	*CYP2D6*
5'exon3_for	ccttcatggccacgcgc	Exon 3	*CYP2D6*
4950r	aagccctgtacttcgatgtcac	Exon 7	*CYP2D6*
cDNA CYP2B6 FE1	atggaactcagcgtcctcctcttc	Exon 1	*CYP2B6*
cDNA CYP2B6 RE7	tctccctgtagactctctctgca	Exon 7	*CYP2B6*
cDNA CYP2B6 FE5	ctgtttgagctcttctctggcttc	Exon 5	*CYP2B6*
cDNA CYP2B6 RE9	tcagcggggcaggaagcggatc	Exon 9	*CYP2B6*
CYP2B6-1F	acattcacttgctcacct	Intron 1	*CYP2B6*
CYP2B6-1R	gtaaataccacttgacca	Intron 2	*CYP2B6*
CYP2B6-2F	atcctactcagaatgatgcacaa	Intron 2	*CYP2B6*
CYP2B6-3R	attacaggtgagagtcatcacatc	Intron 3	*CYP2B6*
CYP2B6-4F	ggtctgcccatctataaac	Intron 3	*CYP2B6*

Primer Designation	5' to 3' Sequence	Primer Location	Gene Specificity
CYP2B6-4R	ctgattcttcacatgtctgcg	Intron 4	*CYP2B6*
CYP2B6-5F	gacagaaggatgagggaggaa	Intron 4	*CYP2B6*
CYP2B6-5R	ctccctctgtctttcattctgt	Intron 5	*CYP2B6*
CYP2B6-8R	tgcaatggttgattgatgctc	Intron 8	*CYP2B6*
CYP2B6-9R	taattttcgataatctcactcctgc	3'UTR	*CYP2B6*
MALDI-TOF MS			
2B6-Tag_Ex4F	acgttggatgtcggtctgcccatctataaa	Intron 3	*CYP2B6*
2B6-Tag_Ex4R	acgttggatgtgattcttcacatgtc	Intron 4	*CYP2B6*
Ex_15582C>T	ggcccctaggaagaagcagca	Intron 3	*CYP2B6*

2.3.6 Software and databases

Product	Supplier
GraphPad Prism Version 4.0	GraphPad Software (San Diego, USA)
HP Chem 4.06	Agilent (Waldbronn, Germany)
Vector NTI Advance 10	Invitrogen (Karlsruhe, Germany)
Ensemble Genbank	http://ensemble.org/
Human CYPallele homepage	http://cypalleles.ki.se
Oligo 6.65	MedProbe (Oslo, Norway)
Scan Array Express 3.0	PerkinElmer (Rodgau – Jügesheim, Germany)
ClustalX (Version 1.81)	freeware (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/)
nucleotide BLAST (blastn)	freeware (http://www.ncbi.nlm.nih.gov/BLAST)
ESE-finder 3.0	freeware (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home)

2.3.7 DNA and liver samples

Blood and liver tissue was previously collected from patients undergoing liver surgery at the Department of General, Visceral and Transplantation Surgery (Prof. Dr. P. Neuhaus and Prof. Dr. A.K. Nuessler, Campus Virchow, University Medical Center Charité, Humboldt University in Berlin, Germany). The preparation of genomic DNA, liver microsomes and protein has been described before (Lang et al., 2001; Lang et al., 2004). Resected tissue samples were morphologically examined and only histological normal liver tissue was used. A total of 235 liver samples were available for protein and activity analysis whereas high quality RNA with a RIN number smaller than 7 was available from a subset of 150 samples (see below). Clinical patient documentation for all samples included age, gender, medical diagnosis, pre-surgical

medication, alcohol use, smoking. None of these samples were from patients with hepatitis, cirrhosis or chronic alcohol use. The study was approved by the local ethics committee of the Charité, Humboldt University Berlin, following the ethical guidelines of the Declaration of Helsinki and written informed consent was obtained from all patients.

2.4 Methods

2.4.1 Probe design for splicing platform

Oligonucleotide probes used for the splicing microarray were chosen to reach a similar hybridization temperature of minimum 71.7°C and maximum 78.2°C. The GC content was less than 45% to minimize secondary structures of oligonucleotide probes. For the *CYP2D6* and *CYP2D7* splicing array 77 probes and for *CYP2B6* 64 probes were designed and analyzed with the Software packages Oligo 6.65 and Vector NTI Advanced 10 package. Only probes showing unlikely self-dimerization and no supposable false priming in the *CYP2D6* and *CYP2B6* gene were chosen. In case of single nucleotide mutations (up to 5 mutations) in the sequence wobble bases were included in the probe structure. In case that there were differences in the highly homologous nucleotide sequence between *CYP2D6* and *CYP2D7* additional probes were designed for the pseudogene. Exon probes, intron n to exon n+1 probes, exon n to intron n probes, exon n to exon n+1 probes (junction probes) and exon n to exon n+2 probes (jump-junction probes) were designed for each exon and intron and placed on the gene structure as shown in figure 4. The *CYP2D6* gene sequences M33388, *CYP2D7* pseudogene sequence M33387 and *CYP2D8P1/P2* gene sequences NG_000854 and NG000853 were used for specific *CYP2D6* and *CYP2D7* probe design. For specific probe design of *CYP2B6* the sequence NG_000008 which also included the pseudogene *CYP2B7* was used. Three control oligonucleotides were designed. Their sequence origin is *Arabidopsis thaliana*. The positive hybridization control (HybrPr) is a probe to control hybridization events. A complementary Cy3-labelled probe is added during hybridization. Negative hybridization probe (NegHybrPr) is a probe, which shows no hybridization result. The spotting probe is labelled with Cy3 on the 3'-side and is added to control the spotting process. All oligonucleotide probes were desalted purity grade and MALDI-TOF MS quality checked. The sequences of the probes are given in the appendix/ table 1.

2.4.2 Optimisation of spotting conditions

For initial spot test 20 pmol of a selection of oligonucleotide probes including the HybrPr were spotted on epoxy coated glass slides (Slide E, Nexterion) testing different spotting buffers. The following buffers were used: 3xSSC, 3xSSC+1.5 M

betaine, 2x spotting solution X and 2x spotting buffer Y (2.2.1). Using spotting buffer Y the influence of spot size and intensity was analysed with following concentrations of oligonucleotide probes: 20 pmol, 15 pmol, 10 pmol and 5 pmol.

2.4.3 Fabrication of the splicing array

All 141 oligonucleotide probes (appendix/ table 1) were spotted on epoxy coated glass slides (Slide E, Nexterion) in triplicate. On the edge of each subgrid three control probes were spotted: a positive hybridization control (HybrPr) complementary to a labelled oligonucleotide target (5'- Cy3-gctcctgactcgtccaatc), which was spiked in while hybridization, a negative hybridization probe (NegHybPr) and a Cy3- labelled spotting probe (SpotPr). For contact printing of presynthesised probes on glass surface a Microgrid II (BioRobotics) high throughput printer and wolfram pins (MicroSpot 2500 pins, BioRobotics) with a capillary slot were used. The probes were dissolved in 2x spotting buffer Y to a final concentration of 20 pmol and soaked up by the submerged pin by capillary forces. The force generated from the downward movement and the surface tension on the slide draws amount of the solution out of the split (Cartegni *et al.*, 2003). For optimal printing results the spotting parameters involving spotting buffer, surrounding humidity, slide surface, probe length and pins had to be optimized empirically. For covalent immobilization, the oligonucleotide arrays were incubated at 60°C for 30 min in a drying compartment (Memmert). For blocking, the slides were rinsed for 5 min in rinsing solution 1.5 min in rinsing solution 2, and subsequently 10 min in rinsing solution 3 with constant stirring. Then, the slides were incubated with the spotted site up in blocking solution at 50°C in a heating compartment (OV5, Biometra). To clean the slides, the slides were rinsed in ddH_2O for 1 min and then dried under a flow of nitrogen.

2.4.4 Target preparation

For hybridization Cy3 labelled DNA was used for validation of the microarray probes and Cy3 labelled cRNA was prepared for detection of alternative splicing.

2.4.4.1 Cy3-dCTP labelled target DNA preparation

A previously described cDNA vector pBS SK+2D6 expression vector (Evert *et al.*, 1997) and a genomic *CYP2D6* pCMV4 expression vector (Raimundo et al., 2004; Toscano et al., 2006) was used to amplify PCR fragments which were labelled and used to evaluate the specificity of the used oligonucleotide probes on the developed splicing microarray for *CYP2D6* and *CYP2D7*. Using the *CYP2D6* cDNA pCMV4 expression vector a fragment comprising exon 1-9 was amplified using the forward primer 2D6_Ex1_Maldi and reverse primer 2D6_rev3_Maldi. Three different fragments including intronic regions were amplified using the genomic *CYP2D6* pCMV4 expression vector as target DNA. The first fragment (exon 1 to intron 4) was amplified using the forward primer 2D6_Ex1_Maldi and the reverse primer 5'exon3_rev, the second fragment comprising intron 3 to exon 9 was amplified using forward primer 5'exon3_for and the reverse primer 2D6_rev3_Maldi and a third fragment comprising intron 3 to exon 7 was amplified using the forward primer 5'exon3_for and the reverse primer 4950r.

For evaluation of *CYP2B6* probes the cDNA vector pIKAT15 was used for amplification of fragments including only exonic regions. Genomic DNA (with genotype *1/*1) was used for amplification of exonic and intronic regions. Six short fragments including intronic regions were amplified labelled and hybridized to evaluate the intron-exon probes: exon1 to intron 1 (primers: CYP2B6-1F and CYP2B6-1R), intron 1-intron 3 (primers: CYP2B6-2F and CYP2B6R), intron 3-intron 4 (primers: CYP2B6-4F and CYP2B64R), intron 4-intron 5 (primers: CYP2B6-5F and CYP2B6-5R), intron 5-intron 8 (primers: CYP2B6-5F and CYP2B6-8R), intron 8-exon 9 (primers: CYP2B6-9F-CYP2B6-9R).

For amplification the Expand High Fidelity PCR system was used (Roche Diagnostics). Reaction was carried out in a 25 µl reaction containing 1.5 mM $MgCl_2$, 200 µM of each dNTP, and 2.6 U of Taq Polymerase. PCR cycling conditions included an initial step at 94°C for 5 min, denaturing at 94°C for 30 sec, annealing at 61°C for 30 sec, extension at 70°C for 1 min for 40 cycles and a final extension at 70°C for 7 min For generating labelled target DNA for microarray analysis, dNTPs including fluorescently-labelled Cy3-dCTP (Amersham Bioscience) were used for PCR (50 µM dATP, dGTP, dTTP, 25 µM dCTP, 25 µM Cy3-dCTP). Amplification was performed in a gradient cycler Ptc-225 (MJ Research). The PCR product was purified

with Qiaquick Spin PCR purification kit (Qiagen) according to the manufacturer's protocol. The DNA was eluted in 30 µl of ddH_2O. The incorporation rate of Cy3-dCTP, expressed as number of nucleotides divided by the incorporated fluorescent dye (NT/F) was determined by OD measurement and was between 70 – 200 Nucleotides/Cy3 (ND-1000 Spectrophotometer, NanoDrop Technologies).

2.4.4.2 cRNA preparation from liver RNA

For cRNA target preparation a total of 5 µg of total RNA origin from liver RNA were reverse transcribed at 50°C using 200 U of Superscript III and 100 pmol of one of the specific primers for either *CYP2D6*, *CYP2D7* or *CYP2B6* which include the T7 promoter region placed in the 5'-end of exon 9 (CYP2D6-T7, CYP2D7-T7 or CYP2B6-T7). Subsequently second strand DNA synthesis was performed at 16°C for 2 h in a final volume of 150 µl adding 30 µl 5x second strand buffer, 40 U DNA Polymerase I, 20 U RNase H and 10 U DNA ligase. Afterwards the second strand DNA synthesis was terminated by adding 30 U of T4-Ligase and incubated 10 min at 16°C.

Now the dsDNA was purified using a Qiaquick Spin PCR purification kit (Qiagen) according to the manufacturer's protocol. The eluted volume was reduced to 14 µl using a vacuum concentration system and afterwards the total volume was reverse transcribed using a MEGAscript T7 transcription kit (Ambion) with aminoallyl labelled UTP (aaUTP). Aminoallyl-cRNA was purified using an RNeasy mini purification kit (Qiagen). The cRNA was eluted with 60 µl of RNase–free water (Ambion). Subsequently the purified cRNA was concentrated to a few µl and 14 µl of labelling buffer and a Cy3-post-labelling reactive dye pack (GE healthcare) diluted in 14 µl DMSO was added. Labelling reaction was performed for 90 minutes at 25°C and was terminated using 5 µl of termination solution (4M hydroxylamine). The labelled cRNA was purified using an RNeasy purification kit (Qiagen). After each step the DNA or cRNA concentration was monitored using the ND-1000 Spectrophotometer (NanoDrop Technologies). The total RNA before starting reverse transcription and the labelled cRNA after T7 transcription were quality controlled using the Agilent Bioanalyzer 2100.

2.4.5 Hybridization

Hybridization protocols were optimised and used for target DNA and for cRNA. For optimal hybridization conditions fragmentation of Cy3-labelled DNA was optimised using 1, 5, 10, 15, 50mU/ µl of DNase and in case of labelled cRNA using fragmentation buffer (Ambion) with varying the fragmentation time between 1 and 15 minutes. Hybridization temperature was optimised between 50-70°C and concentration of hybridization buffer varying between 2x and 6xSSPE or 2xSSC buffer, addition of SDS 0.001 - 0.1% or formamide 0 - 40%. Additionally the wash protocols, using the automated hybridization station, were optimised.

2.4.5.1 Hybridization of DNA

In case of Cy3 labelled DNA, 75 ng of purified PCR product was cut to achieve a fragment length between 100 - 250 bp. For this 0.34 U of DNase (Promega) were used in a volume of 20 µl which was incubated 5 min at room temperature. Afterwards 3 µl of stop buffer (Promega) was added and incubated for 10 min at 60°C. This was controlled with the Agilent 2100 Bioanalyzer. After addition of 0.05 pmol labelled control DNA (5'-Cy3-gctcctgactcgtccaatc) 25 ng of fragmented target DNA was hybridized under a 15 x 16 mm gene frame and cover slip in 70 µl of 2x SSPE in addition of 0.01% of SDS. For hybridization of the DNA – the solution was hybridized under a gene frame and was incubated in a thermomixer with an exchangeable Thermomixer Slide Adapter for 3 hours at 1200 rpm and 70°C. After hybridization the slides were washed with washing solution I, II and III each time 10 min at room temperature, with agitation in a glass container. The slides were then dried with N_2.

2.4.5.2 Hybridization of cRNA

In case of hybridizing labelled cRNA 1 - 5 µg were fragmented for 5 min at 70°C with RNA fragmentation buffer (Ambion) and denatured at 95°C for 10 min. After adding 5 fmol/ µl labelled control DNA (5'-Cy3-gctcctgactcgtccaatc) the fragmented cRNA was hybridized in a volume of 80 µl with 6% SDS and 1.75x SSPE in an automated hybridization station (Tecan400; Germany). Using this machine the glass slides were soaked with 2xSSPE at 23°C. Prior hybridization the temperature was increased to

64°C and the target was added. Hybridization time was 18 h at 64°C and hybridizations were always performed in duplicate on two glass slides in parallel. Washes were performed twice with washing buffer 1 at 23°C for 4 min, then with the same buffer twice at 23°C for 2 min and twice for 4 min in washing buffer 2. Finally the slides were automatically dried with N_2.

2.4.6 Microarray analysis

Microarray analysis consists of two steps first image acquisition and processing and secondly quantitative processing of the microarray results.

2.4.6.1 Image acquisition and processing

After hybridization and washing of the microarrays data from the arrays were extracted by acquisition of the fluorescence signals with a Scan Array Express Scanner (Perkin Elmer). Image processing and calculation of the signal intensities were performed with Scan Array Express Software (Perkin Elmer). For calculation of the individual net signal intensities, the local median background was substracted from the mean raw spot intensity. Further processing of the data was performed using Microsoft Excel software (Microsoft). The mean signal intensity of 7 array control probes (Ambion) nonspecific for human DNA sequence and spotted in triplicate was substracted as mean unspecific background. Finally the mean of the three spots of each probe and the mean of the duplicate experiments were calculated. These values were used for further analyzing the results.

2.4.6.2 Quantification process of microarray results

For analysis of aberrant splicing after microarray scanning, image processing and background determination, the relative intensities of exon-intron, intron-exon, junction and jump-junction probes were calculated. Probes, which overlap in their sequences, were compared. The highest signal in each group was assumed to represent the most prevalent mRNA transcript and was set at 1. From all other probes in this group the relative intensities based on the probe with the highest intensity were calculated. Transcript diversity was calculated by setting the amount of all transcripts, including signals for junction, jump-junction, exon-intron and intron-exon, at 100%.

2.4.7 *CYP2B6* genotyping using MALDI-TOF MS

The intron mutation 15582C>T was genotyped in a collection of 235 liver samples using a previous in house optimised genotyping system (Blievernicht et al., 2007) based on primer extension and mass detection via MALDI-TOF MS. In brief, 50 ng of whole genomic DNA was prepared from whole blood using a QIAamp DNA blood kit (Qiagen GmbH, Hilden, Germany). The genomic DNA was used as a template for amplifying a fragment of 526 bp with a forward primer placed in Intron 3 2B6-Tag_Ex4F 5'- and a reverse primer in Intron 4 2B6-Tag_Ex4R. The PCR reaction was performed in a final volume of 8 µl and contained 0.2 µl units of HotStarTaq polymerase, 1 mM $MgCl_2$ (both Qiagen GmbH) and 0.1 µM of both primers. PCR conditions were 95°C for 15 min followed by 45 cycles at 95°C for 30 s, 60°C for 1 min, and 72°C for 1 min; and finally 72°C for 10 min. The used oligonucleotides were specific to discriminate between *CYP2B6* and the pseudogene *CYP2B7*. *CYP2B6*-specific amplification was verified by sequence analysis.

Following the PCR reaction, residual dNTPs were dephosphorylated using 0.3 U of shrimp alkaline phosphatase in 1xRx-Buffer (Amersham, Freiburg, Germany) for 20 minutes at 37°C and 10 minuites at 85°C. The PCR products were used as templates for primer extension reactions. The extension reaction with a final volume of 16 µl contained 1 µl of buffer C (500 mM Tris HCL, pH 9.5; 100 mM $MgCl_2$), 6 mM $MgCl_2$ 1U of Termipol DNA Polymerase I (all from Solis Biodyne, Tartu, Estonia), 0.12 mM dNTPs and ddNTPs (dATP, dGTP; Amersham and ddC, ddT; BioLog Life Science Institute, Bremen, Germany) and 1 µM of extension primer (Ex15582C>T). Extension reaction were performed at 94°C for 4 min followed by 55 cycles at 94°C for 30 s and 60°C for 30 s and 72°C for 30 sec; and finally 72°C for 2 min. The final base extension products were treated with a cationic exchange resin (AG® 50W-X8 Resin; Bio-Rad Laboratories, Inc.) for 30 min to remove salts. Reaction solutions (0.5 µl) were dispensed onto the 384 well format Anchor ChipTM target plate (Bruker Daltonik, Leipzig, Germany) prespotted with 1 µl matrix solution of 3-hydroxypicolinic acid. Mass spectrometric analyses were performed using an Ultraflex Mass Spectrometer (Bruker Daltonik). The extension primer itself has a molar mass of 6865, in case of ddT or ddG extension the molar mass was increased to either 7466.91 or 7178.79 respectively. The results were analysed using the Genotools software (Bruker Daltonik).

2.4.8 Analysis of genotyping data

Liver samples had been previously genotyped for the most important *CYP2B6* SNPs or respective alleles as described (Blievernicht et al., 2007). Data of the previously genotyped liver samples were used and analysed using Prism 4.0 for explanation of genotype dependent variations and analysis of alternative splicing.

2.4.9 Preparation of RNA and reverse transcription (RT)

Total RNA from liver tissue or cells was isolated using either the RNeasy MidiKit (Qiagen) or Trizol reagent with subsequent RNA clean-up using the RNeasy mini kit (Qiagen) with on column DNase I treatment according to the suppliers instructions. RNA quality was controlled by Agilent Bioanalyzer using RNA6000 Nano Lab Chip Kit (Agilent Technologies, Waldbronn, Germany). Only RNA samples with a RIN number higher than seven (n=150) were used for further analyses. Each sample was reverse transcribed using 1 µg RNA, random hexamers and the TaqMan Reverse Transcription Kit (Applied Biosystems, Foster City, CA). Reverse transcription of RNA for subsequent TaqMan analysis was carried out at 25°C for 10 min, at 48°C for 30 min, and at 95°C for 5 min. For alternative splicing analysis the second step was elongated to 90 min in order to achieve full-length cDNA transcription. After RT-PCR agarose gel and the Agilent Bioanalyzer (Agilent Technologies) were used to analyse and quantify the splicing products.

2.4.10 Quantitative real – time RT-PCR

Quantification of *CYP2B6* mRNA was performed by specific quantitative TaqMan real-time RT-PCR assays on an ABI Prism 7500 system (Applied Biosystems), as described previously (Zukunft et al., 2005). Primers (TQ-CYP2B6for/ rev) for CYP2B6 and 5-carboxyflourescein labelled probe CYB2B6MGB was located on exons 4 and 5 (for primer sequences see appendix, table 1 and figure 35) and were used at concentrations of 400 nM. PCR was performed using 2x universal PCR master mix (Applied Biosystems) in a final volume of 25 µl and the following cycle conditions: 50°C for min, 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min For normalization a predeveloped ß-actin TaqMan assay was used (Applied

Biosystems). The full length *CYP2B6*-cDNA was used as positive control and the splicing variant plasmids lacking exon 4-6 (SV1) or only exon 4 (SV8) were used as negative controls.

2.4.11 Cloning and sequencing of alternative splicing products, RT-PCR

Two *CYP2B6* cDNA fragments spanning exon 2/3 to 9 were generated by RT-PCR. Primers used for the first fragment (764 bp) from exon 2/3 to 7 were cDNA2B6FE2 and cDNA2B6RE7. The second fragment from exon 5 to 9 (931 bp) was amplified with the primers cDNACYP2B6FE5 and cDNACYP2B6RE9. The PCR reaction (final volume, 25 µl) 2 mM dNTPs, 2.5 U of TaqPolymerase, 1x PCR buffer and 7 pmol of each primer was cycled using following conditions 94°C for 2 min, 30 cycles with 94°C for 1 min, 60°C for 30 s, 72°C for 1 min; and finally 72°C for 10 minutes. PCR products were analysed using a 1.5% agarose gel stained with ethidium bromide. Fragments were cut out and extracted using a Gel Extraction kit (Qiagen).

2.4.12 Cloning of RT-PCR fragments

Fragments were cloned into pCR4-TOPO vector and transformed in TOP10 chemically competent cells (both Invitrogen, Karlsruhe, Germany) using the described protocol from Invitrogen. Highly pure plasmid was obtained by means of Qiagen Mini/MidiPrep (Qiagen). Concentration of the plasmids were determined using the NanoDrop device.

2.4.13 Sequencing

5-15 µl of pure plasmid was sequenced with universal primers M13 forward and reverse and the previously described *CYP2B6* sequencing primers were used (Lang et al., 2001), as shown in 2.3.5). Sequencing reaction was performed in a final volume of 21.5 µl containing 15 µl water (including target DNA), 5 µl of primer mix (2 µl DMSO, 1 µl H_2O, 2 µl sequencing primer (2 pmol/ µl)) and 1.5 µl of master mix (Thermo Sequenase Fluorescent Labelled Primer Sequencing Kit with 7-Deaza-dGTP with either didesoxynucleotide ddATP, ddCTP, ddGTP or ddTTP). Before starting the cycling reaction one drop of Chill Out 14 Liquid Wax (Biozym) was

added. Reaction was cycled with following conditions 95°C 2 min, 29 cycles with 95°C 15 s, 57°C for 30 s and 70°C for 1 min. After cycling the reaction was terminated by adding 5 µl of sequencing loading dye. Sequencing was performed using the 4300 DNA Analysis System (LiCor Biosciences). The results were analysed using the eSeq Software (LiCor Biosciences) and the Vector NTI 10.3 (Invitrogen).

2.4.14 Construction of *CYP2B6* minigenes/ expression clones including intronic regions

A minigene including intronic regions and the reading frame from the ATG transcription start site to TAG stop was designed using Vector NTI 10.3 (Invitrogen). Two overlapping fragments 1 (3073 bp, comprising exons 2 to 4; primers 2B6_3'ecoNItaq_E2_for1 including a 5' *Eco*NI restriction site and 2B6_5'sacItaq_E4_rev1 including a *Sac*I restriction site) and fragment 2 (3877 bp, comprising exons 5 to 8; primers CYP2B6-5F and 2B6_spez_I8R_rev1) were amplified from *CYP2B6*1/*1* genomic DNA samples (primer sequences in 2.3.5). Following ligation into a pCR-4TOPO vector (Invitrogen) in *E.coli* Top10 chemically competent cells (Invitrogen) clones showing reference sequence NG_000008 were selected by sequencing using previously described primers (Lang et al., 2001). An internal fragment of 1437 bp was excised from intron 3 using the restriction enzyme *Mfe*I. Intron 6 was also shortened by a fragment of 1535 bp using enzyme *Bsg*I. A previously described *CYP2B6* cDNA vector (pIKAT15; (Lang et al., 2001)) was modified by digestion with *Hind*III and *Sbf*I and replacing the insert by a 518 bp fragment generated by the primers 2B6_5'SbfI_Ex7_for1 and 2B6_spez_I8R_rev1 to include an additional *Eco*NI restriction site by the reverse primer (figure 25 C, D). Both fragments 1 and 2 were digested with *Eco*NI and *Sac*I restriction enzymes and were combined with *Eco*NI-digested modified pIKAT15 in a three fragment ligation reaction (figure 25 B, C).

The resulting parental minigene construct [*1] was confirmed by sequencing. Single mutations were introduced by in vitro mutagenesis into the parental [*1] construct using the QuickChange site directed mutagenesis kit (Stratagene, Amsterdam, Netherlands). Primers used were 2B6_Ex4_15631T_for/rev (15631G>T in exon 4), 2B6_Ex5_777C_for/rev (18045C>A in exon 5), and 2B6_Ex5_18053G_for/rev (18053A>G in exon 5; see appendix table 1). Successful mutagenesis was confirmed

by sequencing and fragments obtained by sequential restriction digestion with enzymes *Mlu*I and *Acc*651 or *Sbf*I and *Acc*651 were subcloned into the wild-type [*1] vector (figure 25 and 26). The resulting minigenes were designated by indicating the allele number set in brackets to indicate the fact that they harbour only the respective key-mutation(s), e.g. minigene [*6] equals the construct with both SNPs c.516G>T and c.785A>G.

2.4.15 Expression in COS-1 and Huh7 cells

COS-1 cells were cultivated in modified DMEM medium (2.3.2). Huh7 cells were cultured with MEM medium (2.3.2). Both cell lines were cultured at 37°C in a 5% CO_2 atmosphere. On the day before transfection cells were seeded into 12-well plates ($2.5x10^5$ cells per well with 1 ml medium). One hour before transfection culture medium was changed. FugeneHD was used to transfect the minigene-vectors according to the manufacturers instructions. The transfection mix included 1.5 µg of the respective expression plasmid and 300 ng pCMVβ. Additional transfection experiments were performed reducing the amount of minigene construct-DNA by dilution (1000-fold) with empty vector.

2.4.16 Protein preparation of cells for western blot and activity measurements

For protein preparation for western blot, cells were scraped off the wells in PBS and were centrifuged for 1 minute at 10000 rpm. Cell pellets were resuspended in 250 µl lysis buffer (2.2.3) and incubated for 30 minutes on ice. Afterwards the cell-lysate was centrifuged for 2 min at 14000 rpm. Supernatant was used for immunoblotting and determination of protein content using the Bradford method (Bio-Rad). For the isolation of protein used for activity measurements, cells (from two wells) were disrupted using lysing Matrix D and a Fast Prep System (both from Qbiogene Inc) two times at level 6 for 10 seconds. Cell homogenates were cleared by centrifugation for 1 min at 14000 rpm.

2.4.17 Immunoblotting

Data from CYP2B6 apoprotein quantification in liver microsomes (Desta et al., 2007; Lang et al., 2001) was used for genotype dependent analysis. The same method as used in the two paper mentioned above was used for quantification of CYP2B6 apoprotein 48h after cell lines were transfected with *CYP2B6* cDNA expression vectors or *CYP2B6* minigene constructs. CYP2B6 apoprotein was quantified using Western blotting with lymphoblast recombinant Cytochrome P450 2B6 (BD Biosciences) as standard protein. Samples were separated on 10% SDS polyacrylamide gels and blotted on nitrocellulose membranes. The transfer was controlled by Ponceau S staining. CYP2B6 protein was immunodetected with a specific monoclonal mouse anti-human CYP2B6 antibody (BD Biosciences). Cells were cotransfected with pCMVß to correct for variability in transfection efficiency. The ß-galactosidase (ß-gal) enzyme from simultaneous transfection was detected on the western blot using anti-ß-gal (MoBiTec, Göttingen, Germany). For control of equal amounts of protein blotted the housekeeping gene ß-actin was detected by using a ß-actin monoclonal antibody (Sigma-Aldrich, Deisenhofen, Germany). In all cases the bound antibody was quantified by secondary IR Dye 800 goat anti-mouse antibody IgG (LiCor Biosciences, Lincoln, NB) using the infrared imaging system 4300 DNA Analyser Odyssey (LiCor Biosciences) with version 1.2 analysis software (LiCoR Biosciences).

2.4.18 Determination of ß-galactosidase activity.

For determination of ß-gal activity 50 µl of cell suspension was mixed with 50 µl of 2x passive lysis buffer (Promega, Mannheim, Germany) and incubated 30 min at room temperature. ß-gal activity was assayed in 10 µl cell extract, which was incubated in 100 µl assay buffer for 30 min. Relative activity of ß-gal activity was determined using Galacto-Light (Applied Biosystems, Darmstadt, Germany) on an Auto-Lumat plus (Berthold Technologies, Bad Wildbad, Germany). 300 µl of ß-gal stop mix was injected automatically and luminescence was measured for 5 s after a delay of 7 s.

2.4.19 CYP2B6 enzyme activity

Bupropion hydroxylation activity was measured in liver microsomes (50 µg) and origin from COS-1 and Huh-7 cell extracts as described previously (Klein et al., 2005; Richter et al., 2005). In brief, 50 µg of protein was incubated in a reaction volume of 0.1 ml. The reaction mixture was equilibrated at 37°C for 3 min and started by adding 10 µl of 10x NADPH-regenerating system (2.2.5). After 15 min at 37°C the reaction was terminated by adding 20 µl 1 M HCl. Substrate concentration was 50 µM for human liver microsomes and 1 mM for CYP2B6 from transfected cell lines. Hydroxybupropion was determined by liquid chromatography-mass spectroscopy as described (Richter et al., 2004). All incubations were performed in duplicate and the data was analysed with Software HP Chem 4.06 (Agilent Technologies)

2.4.20 Allele nomenclature and statistics

Base numbering and allele definitions were according to the published recommendations of the CYPallele Nomenclature Committee (http://www.imm.ki.se/CYPalleles/criteria.htm). Numbering was based on the cDNA with the full-length cDNA sequence published by Yamano et al. and defined as the wild-type (CYP2B6*1) (Yamano *et al.*, 1989). Genotype and allele frequency data are mainly based on the results previously described (Blievernicht et al., 2007). Data were compiled according to the genotype and allele frequencies estimated from the observed numbers of each specific allele and with the assistance of the PHASE program v.2.0.2 (Stephens and Donnelly, 2003; Stephens et al., 2001). Statistical calculations were performed using the GraphPad InStat v3.00 or the GraphPad Prism program (Version 3.03; GraphPad Software Inc., San Diego, CA).

Because some data sets were not normally distributed, nonparametric methods (Mann-Whitney U-test; Spearman rank correlation coefficient r_S) were generally used to compare phenotypic data. Multiple comparisons were performed by one-way analysis of variance and post-hoc testing with Dunnett's test. 95% Confidence Interval of proportions for mutation frequencies were calculated using the Wilson procedure with continuity correction (Newcombe, 1998). All statistical tests were performed two-tailed and statistical significance was defined as $P<0.05$.

3 Results

3.1 Microarray platform – general considerations

As outlined in the introduction, the major problems of analyzing splice variants in human tissues are the large number of in principle possible alternative transcript structures, their very different sizes and amounts expressed, and the existence of highly similar transcripts originating from pseudogenes, i.e. non-functional copies of the genes of interest. The availability of a microarray spotter, a hybridization station and a microarray scanner at the Institute of Technical Biochemistry Stuttgart made it possible to adress these problems with a self-designed microarray. Based on these difficulties, the following challenges were envisioned for a self-designed splicing microarray: i) a systematic probe-design that would allow to detect most or all transcript variants expressed at significant level; ii) a way to permit the distinction between gene- and pseudogene-specific transcripts; iii) the design should be tolerable against polymorphic variation within the probe sequences; iv) at least semi quantitative data-acquisition in order to allow statistical analysis of relationships between genotype and other host-factors and observed splice variants; and v) a platform easily adaptable to other genes. As a model gene to develop and validate an array design that fulfils these requirements the *CYP2D6* gene was chosen because a large number of reported splicing variants existed and because the expressed pseudogene *CYP2D7* allowed tackling the gene/ pseudogene problem (see table 1).

3.1.1 Probe design

It was first decided to use relatively long probes with a length of at least ~29 - 51 nt. because it was described in the literature that rather long probes would tolerate single nucleotide polymorphisms (Fehlbaum et al., 2005). Toleration of single mutations is inevitable as both genes *CYP2D6* and *CYP2B6* for which the splicing microarray was developed are highly polymorphic. Longer probes were chosen for their uniqueness and as they had been shown to be more sensitive for detection (Castle et al., 2003).

A principal set of probes were designed to detect the most common splicing events (figure 6). The final design included seven different probes of five principal types: for each exon, three internal probes were designed and located in conserved exonic regions (to quantify the mean expression level of each exon); one 5'-exon-intron overlapping probe, one intron-3'-exon overlapping probe (both to detect intron-inclusion); one junction probe stretching from exon n to exon n+1 (to detect the normal transcript) as well as a jump junction probe stretching from exon 1 to exon n+2 (to detect splice variants lacking an exon; figure 6). According to Watahiki et al. (2004) who analyzed the relative representation of different aberrant splicing events within the genome, the five different probes theoretically detect over 70% of all existing splicing aberrations, including exon skipping (38.4%), alternative 3'-prime or 5'-prime sites (22%) and intron retention (11.1%; see figure 2) (Watahiki et al., 2004).

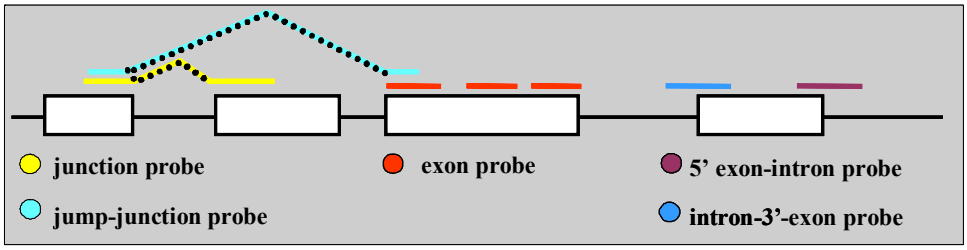


Figure 6: Structure of oligonucleotide probes designed for the developed splicing platform. The figure illustrates the regions where the splicing probes are located within the gene.

In order to cope with the problem of frequent polymorphisms within *CYP2D6* and *CYP2D7*, all available information about known mutations within the sequence were collected and were carefully considered during probe design using the software Oligo 6.65 and Vector NTI Advanced. In case of exon probes regions showing frequent polymorphisms were excluded for probe design if possible. In regions essential for probe location like exon or intron boundaries, up to five wobble bases were included within the oligonucleotide sequence at the polymorphic position.

The gene and pseudogene structure was compared and different probes were designed for both the gene and the pseudogene, whereever possible. In the appendix, in table 1, in the second column, the specificity of each probe shows whether the probe was either used for detection of pseudogene or gene specific transcripts. Due to 97% of sequence homology of the gene *CYP2D6* to the pseudogene *CYP2D7* not for each position a pseudogene specific probe was

possible and where possible the difference was small (only 3 nt). To cope with this problem additionally gene and pseudogene specific primers were designed and used in a gene specific linear amplification Eberwine T7-based protocol. The specific primers included a T7 promoter on the 3'side and a 5'gene specific region situated in exon 9 of the investigated genes *CYP2D6*, *CYP2D7* and *CYP2B6*.

The optimizations performed by Castle et al. 2003 towards the discrimination capacity of junction probes and optimal length was considered for designing the microarray probes (Castle et al., 2003). The following specifications were fixed in order to facilitate oligonucleotide design and comparison of results: for junction and jump junction probes 50% of the sequence consisted of exon n and 50% of the sequence of exon n+1 or exon n+2; for intron-exon and exon-intron probes around 50% of the sequence consisted of intronic and 50% of exonic region.

As all probes should be used together on one glass slide under the same hybridization conditions the individual probes were adjusted such to show high homogenity. Therefore the probes had to be fine-tuned in order to show high specificity and fit a narrow range of thermodynamic behaviour. During probe design it was specially focused on a narrow range of Tm (mean of all 75.7°C) and a GC-content below 50% to reduce cross hybridization and increase sensitivity (mean of GC content: 37%). All probes were controlled by the software Oligo 6.65 for secondary structures searching the less unfavourable free energy of hairpin bonds (hairpin ΔG), the less unfavourable number of bases forming hairpin bonds, and less free energy of dimers (dimer ΔG). This was done by varying the oligonucleotide length between 29 and 51 nucleotides (mean of all 41 nt) and finding the best fit probe concerning Tm, GC-content and minimal secondary structure. By applying all these guidelines for design of the probes, finally 77 oligonucleotide probes specific for either *CYP2D6* or *CYP2D7* were designed (appendix/ table 1).

During evaluation of the microarray probes with known target DNA (explained below) false positive signals, so called cross hybridization results, were observed if the boundaries of a junction and an exon-intron probe were similar (up to 6 nucleotides) (figure 7A). Sequence similarity can also be observed comparing boundaries of jump-junction, intron-exon or junction probes (figure 7B). Probes, which showed cross hybridization during the evaluation process and sequence similarity at the boundaries, were redesigned. Probes with these characteristics were elongated at one side for 3 to 6 nucleotides (5' or 3'). This increased the specificity in ten

redesigned probes. This is shown for two examples in figure 7. Within appendix/ table 1 the redesigned probes can be identified by the synonym “platte” within the name.

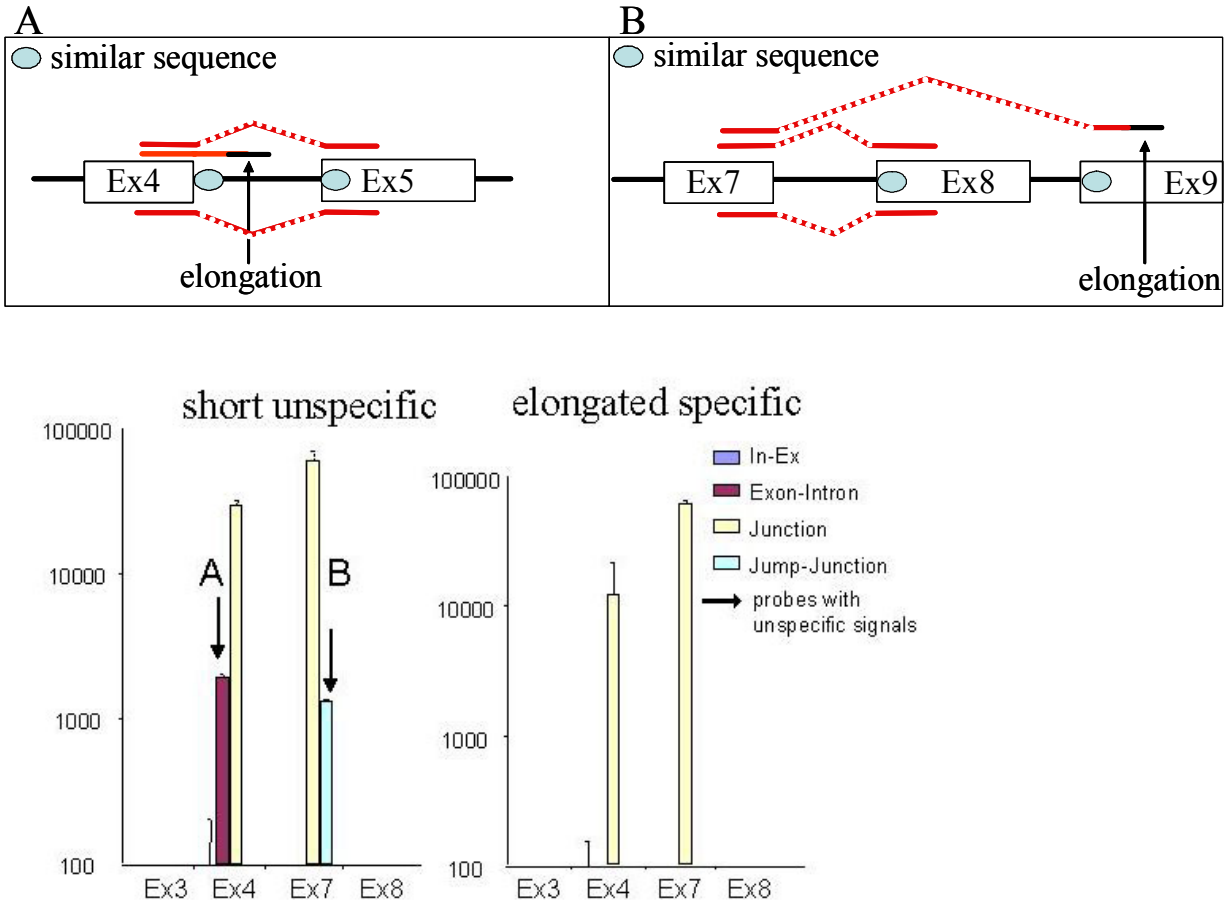


Figure 7: Homology within the boundaries results in unspecific hybridization.

Unspecific signals of hybridization probes could be explained in cases where oligonucleotide probes showed homologies within the boundaries. For example the boundaries of exon 4 to intron 4 and exon 4 to exon 5 showed similarity (redesigned probe: Platte2-A03); also, the boundaries of exon 7 to 9 and of exon 7 to 8 showed similarity (redesigned probe: Platte2-H06).

3.1.2 Array fabrication and layout, spotting tests

The system was set up using epoxy-coated slides. In this system, the spot size, spot uniformity obtained on the microarray slide depend on certain parameters such as the oligonucleotide concentration, the spotting buffer, the humidity during spotting process as well as the pins of the used microarray spotter.

Initial spot-tests and subsequent blocking and hybridizations revealed the dependence of spot size and spotting buffers. An oligonucleotide concentration of 20 pmol/ µl revealed that the smallest spot size (~ 5 µm) was achieved using 3xSSC as spotting buffer, followed by 3xSSC + 1.5 betaine (20 - 40 µm) and spotting solution X (60-80 µm). The biggest spot size and highest intensities were achieved by using spotting solution Y (100 - 120µm). Reduction of oligonucleotide amount (20 pmol/ µl, 15 pmol/ µl, 10 pmol/ µl and 5 pmol/ µl) led to a decrease of spot size (figure 8) and usage of longer nucleotides led to increase spot size (not shown). It was decided to use bigger spot size, as the absolute number of spots is not so high that a limit of space is the consequence. Secondly larger spot size lead to more accuracy as the spot intensity is measured by imaging software by counting the number of pixels and calculating the mean of those compared to the spot size. Thirdly the highest signal to noise ratio was achieved with spotting solution Y. To summarize the most suitable printing buffer for uniform spots in contact printing was the combination of spotting buffer Y together with an oligonucleotide concentration of 20 pmol/ µl.

50% of humidity was maintained during spotting to prevent rapid evaporation of spotted drops, drying up on the microarray pins and efficient coupling of amino-modified oligonucleotides with the epoxy groups on the glass surface. After dipping the pins in the mixture of spotting buffer and probe, the pin was dabbed 20 times on a prespot slide before printing the array for achieving reproducible and uniform spots on the microarray glass slide. Optimised spotting conditions resulted in small standard deviation showing high uniformity and reproducibility of the spotted triplicate spots within one slide and comparing slide to slide (figure 7 and 8). Deduced from the number of pixels (using the conditions mentioned above) the spot size and the spot-to-spot distance were estimated to be 160 µm and 320 µm, respectively (figure 10). Spotted microarrays prepared according to this protocol were stored for a maximum of one month at room temperature, dark and dry until they were blocked and used for hybridization experiments.

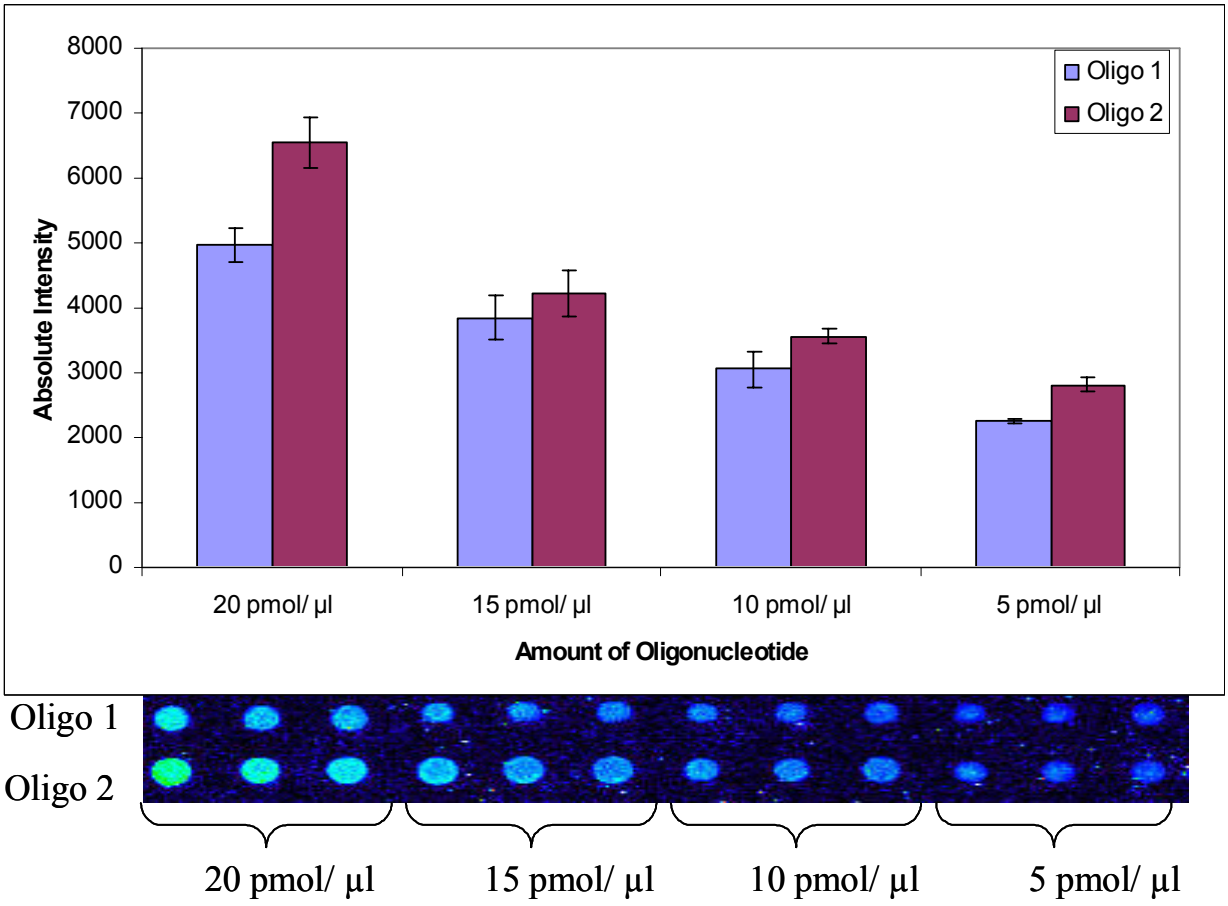


Figure 8: Influence of probe concentration on spot size and hybridization results.
Spotting was performed with different amounts of two oligonucleotides within a fixed volume of spotting buffer Y. Spotting size as well as signal intensity after hybridization increased with higher oligonucleotide concentration.

The layout of the microarray is shown in figure 9. The splicing-probes were spotted in seven columns in the following order: 3'intron-5'exon probe, 3 internal exon-probes, 3'exon-5'intron probe, jump-junction probe and junction probe (figure 9). In every second column (columns on white background, figure 9) pseudogene specific probes were placed. Probes that were exchanged because of discriminatory power were placed within unused space and marked as black dots (figure 9). Within an additional subgrid (not shown in figure 9) seven ArrayControl™ Sense Oligo Spots (Ambion), which were chosen to show no homology to human genes were spotted and used after hybridization as internal background (explained below).

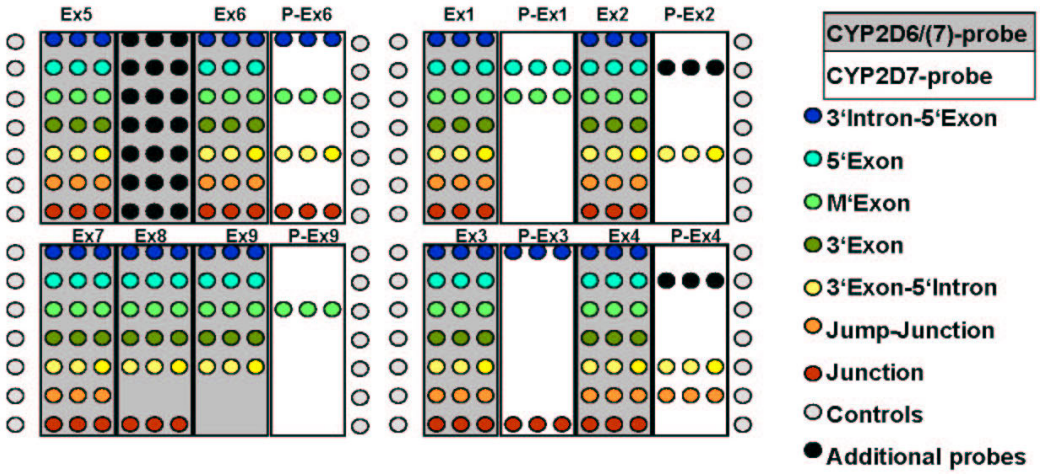


Figure 9: Array layout of the *CYP2D6* splicing array.

The splicing-probes were ordered in seven columns in following order: 3'intron-5'exon probe, three exon-probes, 3'exon-5'intron probe, jump-junction probe and junction probe and marked with a colour code. *CYP2D6* specific probes are shown on grey background, *CYP2D7* pseudogene specific probes are shown on white background. Grey dots are control probes (negative control, positive control and hybridization control).

3.1.3 Optimisation of hybridization conditions using labelled DNA

Initial hybridization experiments were carried out using a thermomixer from Eppendorf with an exchangeable thermomixer slide adapter to find optimal hybridization conditions showing best specificity, discrimination and signal intensity for all oligonucleotide probes. In a first evaluation step Cy3 labelled DNA fragments, amplified from genomic DNA or a vector containing the *CYP2D6* cDNA (pBS SK+2D6) were hybridized to the microarray. Three main components influencing the hybridization result were investigated: the composition of the hybridization buffer, the hybridization time and temperature and the fragment length of the target DNA. Different concentrations of hybridization buffer SSPE (6x, 4x, 2x SPPE, 2xSSC) were tested of which 2x SSPE showed the lowest background and highest stringency (figure 10).

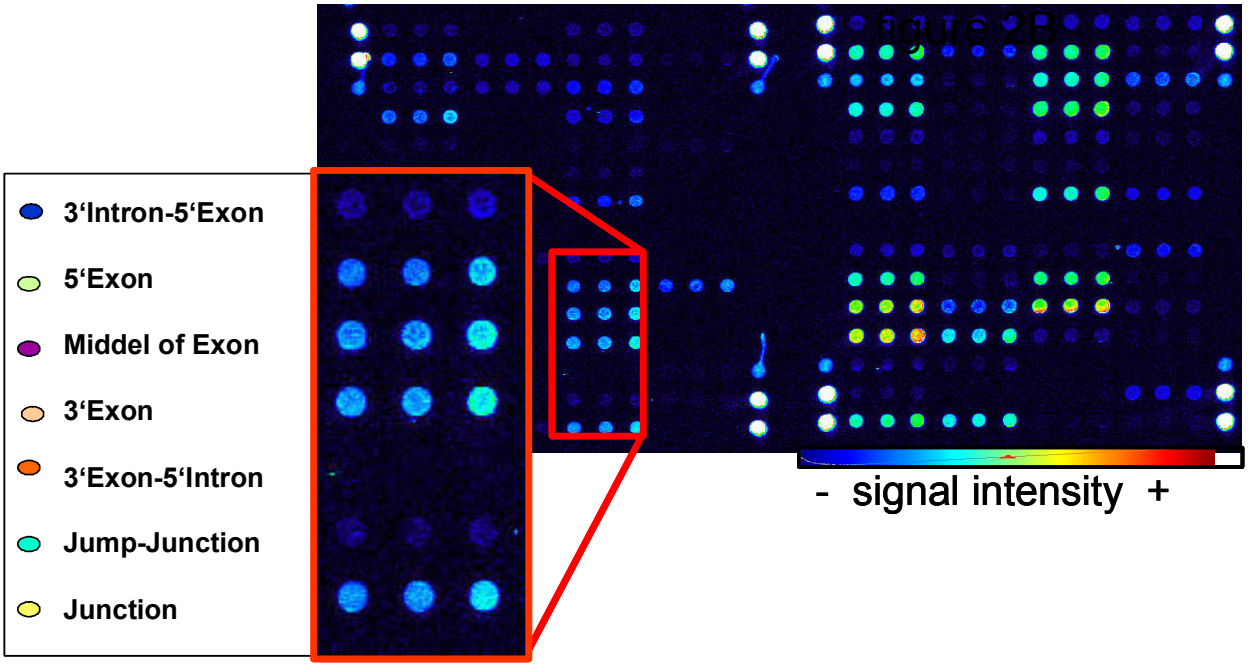


Figure 10: Fluorescent image after hybridization.

Fluorescent image achieved after hybridization of 5 µg target cRNA for 18 h at 64°C using 2xSSPE as hybridization buffer. The signal intensity is shown in false colour. Black corresponds to zero intensity, blue to the lowest signal intensity, red to white depict the highest signal intensities (as output by Scan Array Express 3.0). As shown in the insert on the left all probes were spotted and ordered from top to bottom as described in figure 9

It was observed that the hybridization signal achieved after image acquisition and data quantification decreased for all probes from exon 9 to exon 1 (figure 11). This result was visible after hybridization of labelled DNA fragments amplified with conventional PCR (using forward and reverse primer) and after linear transcription of cRNA (with one *CYP2D6*-T7 primer placed on exon 9).

Analysis of the probe sequences concerning GC content or length revealed no striking difference between the different splicing probes. As different pins spotted exon 1 to exon 4 and exon 5 to exon 9 probes, a possible explanation was that the signal might be reduced by the different behaviour of the spotting pins. But exchange of those revealed the same hybridization result and showed that all 4 pins spotted comparably and uniformly (figure 11). The effect decreased by addition of SDS, fragmentation of the target DNA or cRNA and minimized by analysing the relative intensity of neighbouring probes.

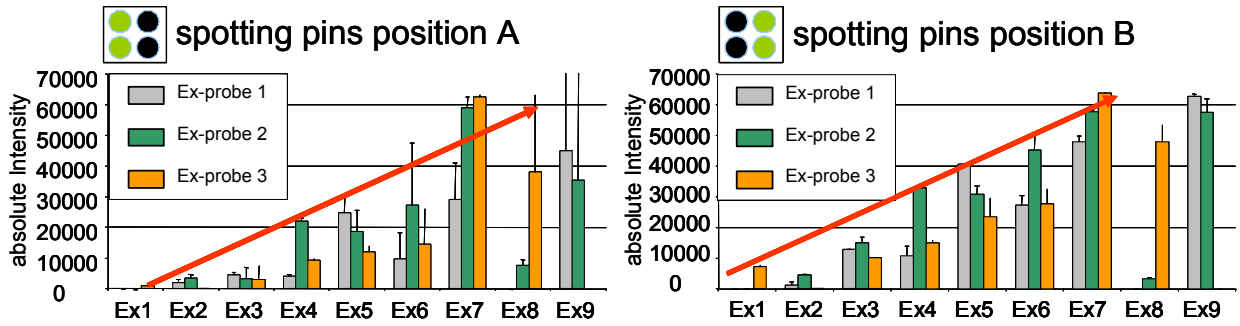


Figure 11: Exchange of left with right pins

Mean net signal intensity is shown for all exon probes after hybridization of 25 ng of Cy3 labelled DNA target including all exons 1-9 and no intronic regions. Exchange of pins resulted in no effect on the signal intensity. The signal increase from exon 1 to exon 9 could thus not be explained by spotting aberrations.

Furthermore, addition of different SDS concentrations (0.001 - 0.1%) to the hybridization buffer was tested. Interestingly, addition of SDS revealed higher signal intensity and 0.1% of SDS showed up to 10 times higher values without showing higher background aberration (figure 12).

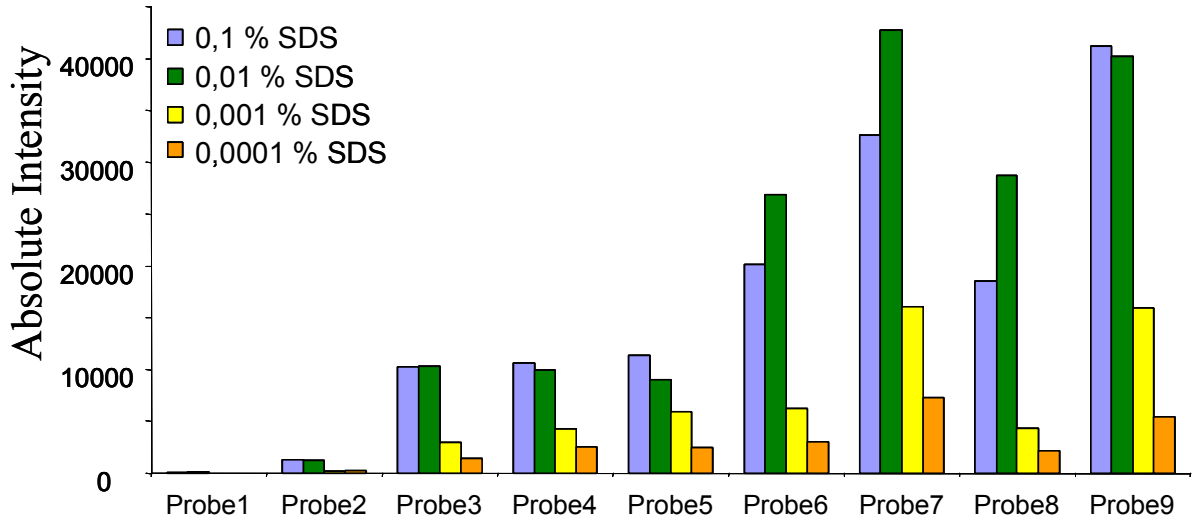


Figure 12: Net signal intensity of nine oligonucleotide probes after background calculation from four hybridization experiments with increasing SDS concentrations and 25 ng of target DNA.

Hybridization temperature for labelled DNA was optimal at 70°C for 3h (2xSSC and addition of 0.1% SDS). By increasing the hybridization time to 18h and decreasing the target DNA stepwise from 50 ng to 1 ng it was shown that labelled DNA was correctly detectable at an amount of only 5 ng. With 1 ng of target DNA no signal was observed for the probes placed within exon 1. To achieve high sensitivity the PCR product had to be cut with DNase to achieve an average fragment length of 100-250 bp. The fragment distribution was controlled with the Agilent Bioanalyzer 2100 as described in 2.4.5.1 and shown for cRNA fragmentation (figure 15).

3.1.4 Data quantification

A new data quantification method had to be developed, as no method existed for this kind of alternate splicing probe set. Before analysing splicing events within human tissue samples, all probes were validated to show correct splicing results in experiments. Either a fragment amplified from cDNA or from genomic DNA was hybridized to simulate either hybridization of conservative spliced transcripts or intronic retention. Only probes showing the expected hybridization result were used (described later). Probes showing a false positive result or no result were redesigned as described in figure 7 or modified by adding the oligo-dT-C7 aminolinke on the 5'site instead of an oligo-dT-C6 aminolinker on the 3' site (indicated in appendix/ table 1). Both methods increased the specificity.

After scanning of the arrays the mean signal of each spot, subtracted by the local background was used for further data calculation. The mean intensity of seven ArrayControl Sense Oligo Spots (Ambion) not specific for *CYP2D6* were calculated on each microarray and subtracted as internal unspecific hybridization signal. After this internal background calculation only signals higher than 100 signal intensity units were used. Any negative intensity values were set as zero.

As described before, hybridizations were always performed in duplicate with each oligonucleotide spotted in triplicate. The mean net signal intensity and the corresponding standard deviation were thus always calculated as mean of the 6 spots obtained in two experiments. This mean signal intensity is herein referred to as the "signal intensity" (I). This way intra- and inter-experiment variability could be calculated. To analyze the transcript diversity always a set of 3-4 probes, which are indicative for a specific splicing event (conservative, or aberrant splicing; figure 13)

were compared. Probes within one probe set can be compared directly as they have sequence homology of 50 percent. They compete for the spliced transcript within this region and the probe being 100% complement to the most common splice product showed highest signal. All others showed weaker signals dependent on the amount of the individual mRNA transcript. Splicing resulted in the most cases in mixed isoforms of different splicing variants.

Two calculation methods were used. In the first analysis method the probe within each probe set showing the highest signal intensity was set as 1, assuming that this probe represents the most prevalent situation in the expressed mRNA. From all other probes in this group the relative intensity (I_{rel}) based on the probe with the highest intensity was calculated (method used in figure 14 and 16).

In the second analysis method all mRNA isoforms are summed up to 100%. All splicing variants are shown in its relative fraction ($I_{fraction/100\%}$) method used in figure 17, 18 and 19)

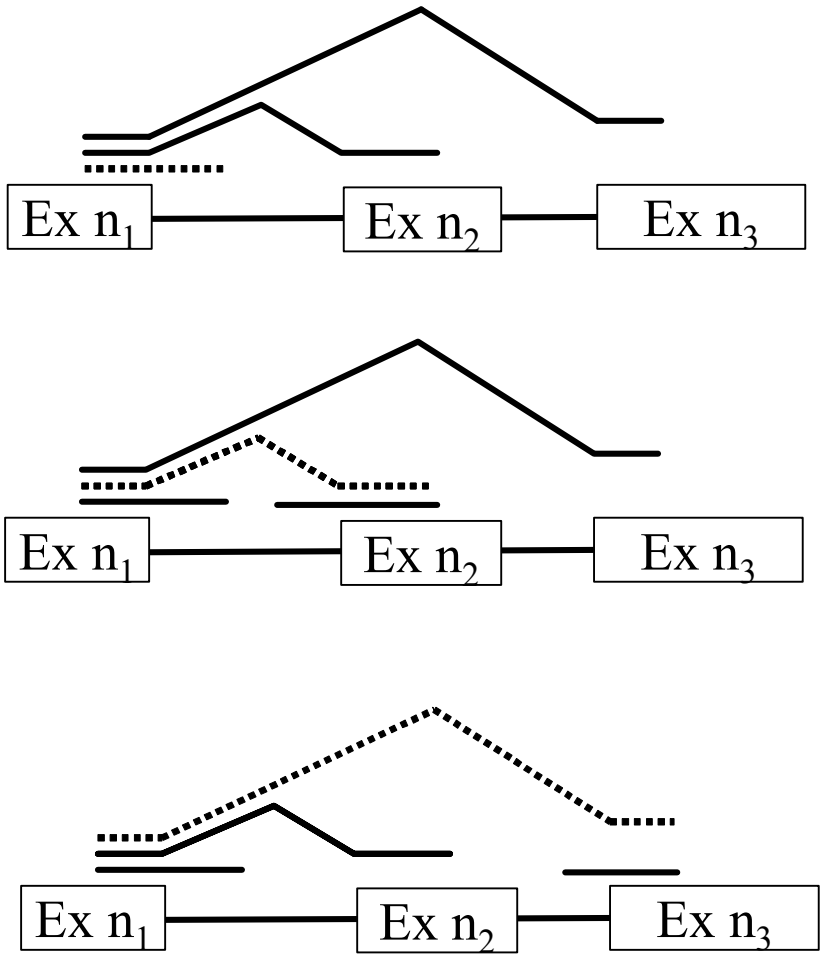


Figure 13: Data quantification

For analysing alternative splicing events the relative intensity was calculated by comparing probes with 50% of sequence homology. Probe sets, which are compared, are shown in this figure. Probes shown in dashed lines are compared to the other probes. Within one probe set, the probe showing the highest intensity was set as 100% and the relative intensities of the other probes were calculated.

3.1.5 Evaluation of probes by hybridization to specific DNA fragments

The microarray probes were evaluated by hybridization to different labelled PCR fragments. One PCR fragment including exons 1-9 was amplified from cDNA vector pBS SK+2D6 and two fragments including intronic regions (encompassing exon 1 to intron 4, and intron 3 to exon 9) were amplified using the existing pCLA 24 vector (containing the *CYP2D6* wildtype genomic DNA) gene expression vector as target DNA (Toscano et al., 2006). Following hybridization of the fragment including all exons, a signal is expected for each exon and each junction probe (to monitor conservative splicing), whereas hybridizing the fragments including intronic regions, signals for intron-exon, exon-intron and exon probes were expected (to monitor intron retention).

In case of hybridizing the labelled cDNA fragment from exon 1-9 all junction probes showed highest signals compared to the other probes (yellow section in figure 14). Unspecific signal were only visible in the case of three probes (intron-exon5, exon5-intron and intron-exon7). These signals were smaller than 15% compared with the highest signal.

Hybridization of PCR products including the intron regions showed high specificity with unspecific signal <5% except for 2 probes with unspecific signals <18% (exon4-exon5 and exon4-exon6 probe, figure 14).

Taken together these experiments demonstrated that the selected probes are highly specific. Five probes (exon4-exon5, exon4-exon6 probe, inron-exon5, exon5-intron and intron-exon7) showed unspecific signals <18%. The evaluated probes were used in further experiments and a cut off was set at 20% compared to the highest signal. All signals smaller than 20% were ignored all signals higher than 20% were analysed as they could indicate alternative transcripts.

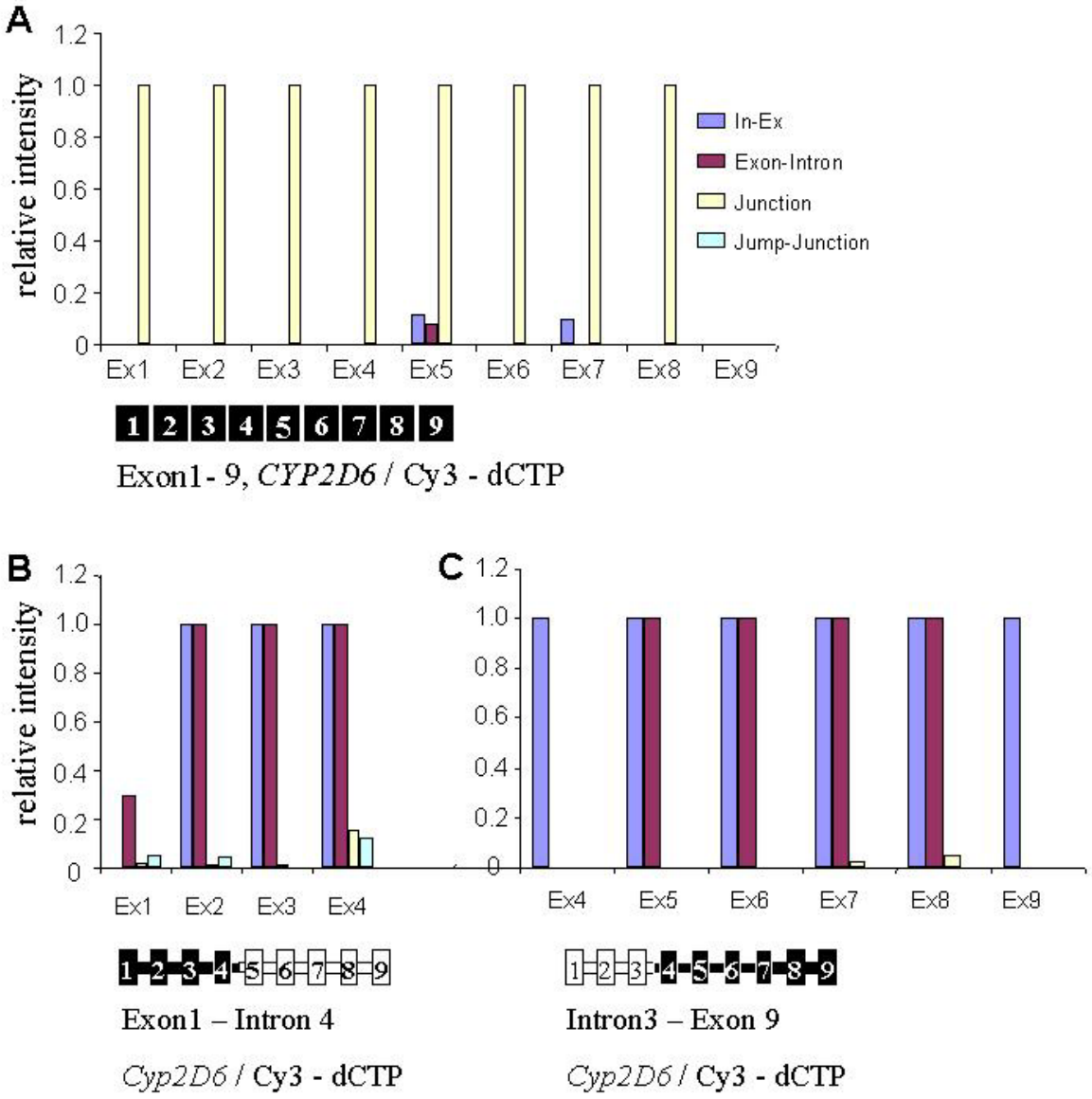


Figure 14: Relative intensities (I_{rel}) of probes after hybridization of 25 ng of target DNA for validation.

A: target DNA spanning exon 1-9 (without intronic region) was hybridized to validate all junction probes. B: PCR product scanning Exon 1 to Intron 4 (including all intron regions). C: intron 3 to exon 9 (including all intron regions) was hybridized to validate all exon-intron and intron-exon probes.

3.1.6 Hybridization results using specific target cRNA

For detection of aberrant splicing in human liver linear amplification of RNA was performed using *CYP2D6* specific primers. One round of linear amplification is an efficient way for target preparation. This result in great integrity of labelled target

cRNA and this way minimize an amplification bias compared to conventional exponential PCR. Starting with 1-5 µg of total RNA up to 25 µg of total mRNA was reached after linear amplification, labelling and all purification steps. The labelling efficiency of cRNA was between 200 and 1000 Nucleotides/ Cy3. This was about 5 times lower compared to labelled PCR product and may be explained by unlabelled total RNA, which was used as target for gene specific linear transcription. After reverse transcription and prior hybridization the length of target cRNA had to be adjusted to 50-250 bp. cRNA was cut with fragmentation buffer (Ambion). The length was optimal between 150-200 nt after 5 minutes incubation time at 70°C (figure 15). Shorter fragmentation time led to longer fragments and at 10 minutes most of the cRNA fragments were shorter than 100 nt. After 35 minutes only fragments between 25-15 nt were obtained (not shown).

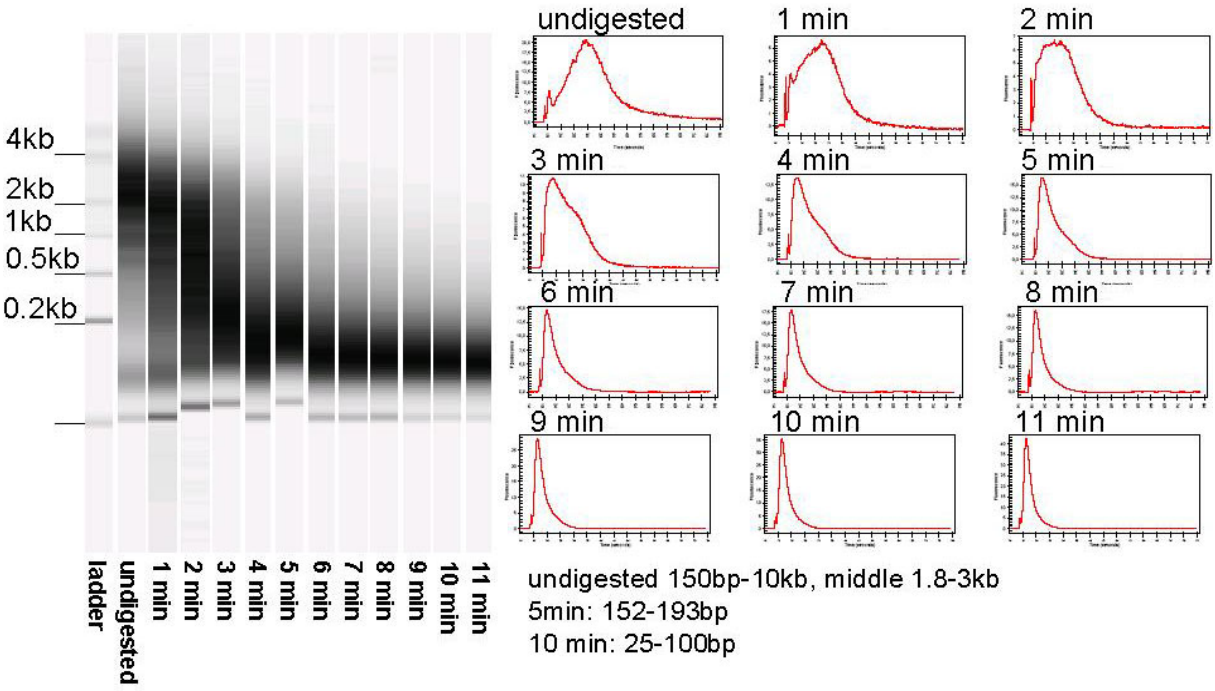


Figure 15: Fragmentation of cRNA

5 µg of cRNA was fragmented and each minute 1 µl was taken out mixed 1:1 (v/v) with 2x stop buffer (Ambion). The diluted cRNA was analysed using the Agilent 2100 Bioanalyzer.

Hybridization experiments with labelled cRNA were always performed in duplicate, using the HS 400 hybridization station and 1 µg -5 µg of cRNA. First hybridization experiments with cRNA using the optimal condition for labelled cDNA resulted in weak signals and high background intensities. Low signal intensities might be due to lower labelling efficiency of cRNA and higher background due to total RNA.

Compared to 75 ng of target DNA used in the hybridization process in case of cRNA an amount of 5-10 µg was needed to achieve signal heights, which were useful for interpretation of the data. Secondly hybridization temperature of 70°C probably led to partial evaporation and this way resulted in high background. To solve this problem optimisation was performed by addition of up to 40% of formamide and reduced hybridization temperature, but this did not solve the problem of weak signal intensity and was discarded. Increasing the hybridization time from 3 to 18 hours and increase of SDS concentration to 7% achieved detectable and analysable signal intensities. Reduction of hybridization temperature from 70°C to 64°C led to decreased background intensity but increased unspecific signals, which was well balanced by increasing the washing steps after hybridization using washing buffer 1 twice for 4 min, then twice for 2 min and afterwards using washing buffer 2 twice for 4 min at 23°C.

3.2 Aberrant splicing in *CYP2D6* and *CYP2D7*

To assess the splicing platform labelled cRNA was prepared and hybridized origin from three individual liver tissues previously genotyped as *CYP2D6*1/*1* (figure 12A) and three samples with genotype **41/*41* (figure 12B). The first analysis was concentrated on exon 5 to 7 as it was known that transcripts lacking exon 6 are enhanced in genotype *CYP2D6*41*.

Hybridizing cRNA carrying *CYP2D6 *1/*1* all junction probes monitoring exon-exon boundaries showed the highest signals (figure 16, A/ yellow columns). For most of the intron-exon and exon-intron probes the signals were below or only slightly above 20 %. Genotype *CYP2D6*1/*1* belongs to the full functional extensive metabolizers. Splicing aberration monitored within this genotype either result from unspecific hybridization or result from the basal level of splicing aberrations, which can be found within every transcriptome and can be categorized as background noise of the splicosome, showing no effect on the phenotype. Therefore splicing aberrants less than 20%, compared to the most common transcript, are ignored.

Hybridized cRNA patterns origin from homozygote *CYP2D6*41*, an intermediate metabolizer, look strongly different from genotype **1/*1*. Within genotype **41/*41* skipping of exon 6 can be seen. The signal of the jump-junction probe (exon 5 - 7) shows the same height as the junction probe (exon 5 - 6; figure 16, B). Interestingly, compared to genotype **1/*1* in **41/*41* the signals for exon 5-intron 5 and intron 5-exon 6 are additionally increased. That means that intron 5 is more often completely or partially retained within **41* homozygotes (figure 12B). Splice variants including the intron 5 or partial fragments of intron 5 are described (Gaedigk et al., 2005) but were so far not assigned to genotype *CYP2D6*41*.

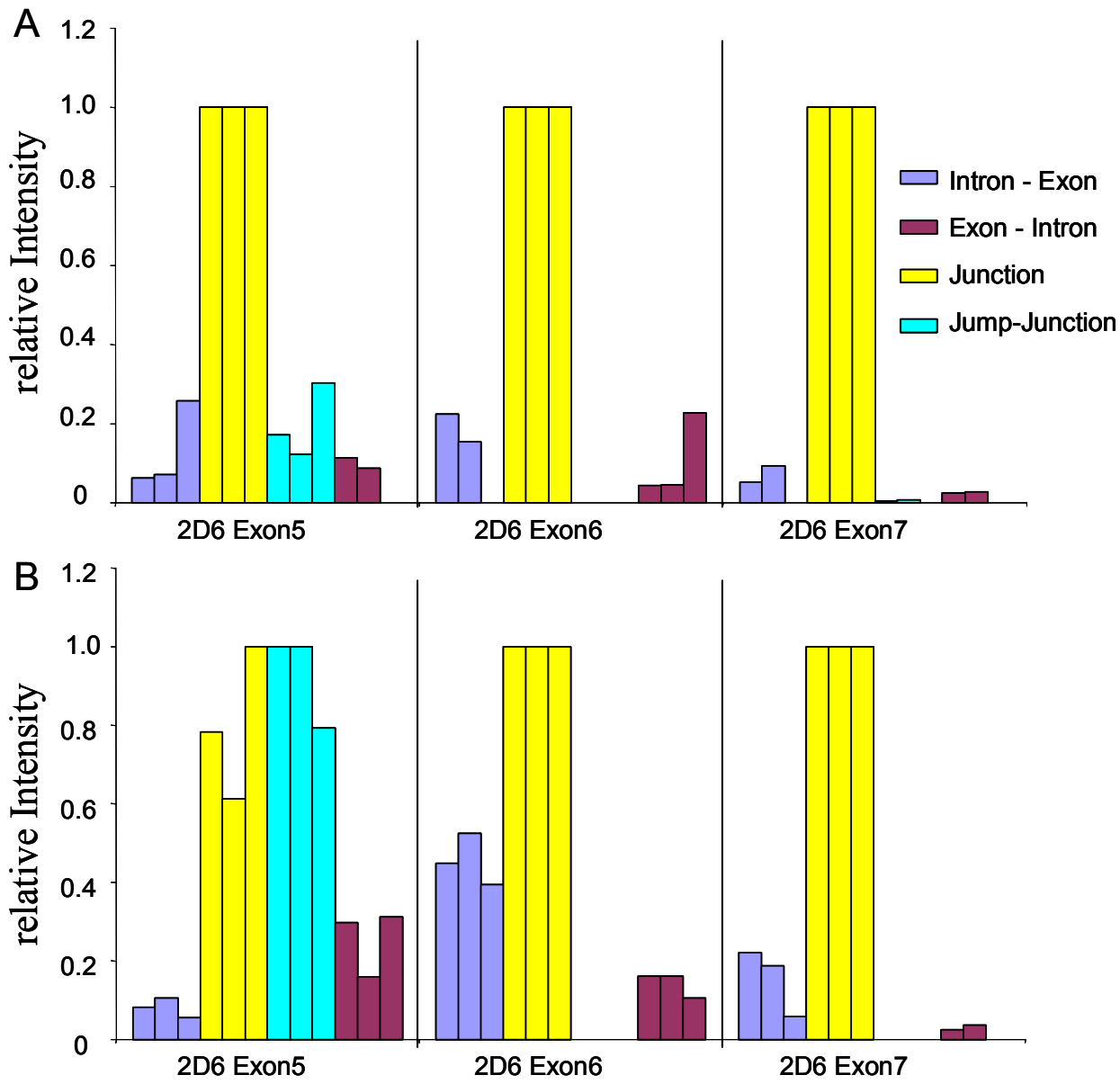


Figure 16: Validation of the microarray using labelled cRNA origin from human liver, which was previously genotyped. 5 µg of linear amplified and labelled target cRNA was hybridized (A) hybridization experiments using 5 µg cRNA originating from 3 individual liver samples with genotype *CYP2D6 *1/*1*. (B) hybridization experiments using 5 µg of cRNA originating from 3 individual liver samples with genotype *CYP2D6 *41/*41*. Relative intensity (I_{rel}) of each probe (spotted in triplicate) was calculated. The yellow bars show the relative signal from the junction probes from exon n to exon n+1. The green bars show the relative signal from the jump-junction probes from exon n to exon n+2. Blue bars show the relative intensity of the intron-exon and the violet bars show the signals of the exon-intron probes.

3.2.1 Gene and pseudogene specific splicing variations

The established platform demonstrated the capability to monitor splicing transcripts using biological material. To find out if there exist additional splicing variants, which can be assigned to genotypes, exons 1-9 were analysed. The splicing array was capable to discriminate between pseudogene and gene specific splicing aberrations. Splicing variants originating from the highly homologous pseudogenes were currently overseen by the previously published splicing microarrays (Brinkman, 2004; Clark et al., 2002; Fehlbaum et al., 2005; Hu et al., 2001; Johnson et al., 2003; Modrek et al., 2001): For discrimination of splicing patterns between gene and pseudogene total RNA was linear reverse transcribed using a gene specific primer for either *CYP2D6* or *CYP2D7* placed on the 3'site of the gene. 15 samples of total RNA originating from human liver tissue were reverse transcribed for *CYP2D6* and *CYP2D7* respectively. In 2 samples (with genotype *CYP2D6*1/*1*) the obtained amount of reverse transcribed cRNA concentration of *CYP2D7* was too low for hybridization experiments. In case of *CYP2D6* as well of the pseudogene *CYP2D7* 1-5 µg of labelled cRNA was hybridized in duplicate on the microarray.
With the developed splicing platform it was possible to monitore allele specific splicing patterns (figure 17). In figure 17 on the left side the mean transcript diversity of 4 human liver samples carrying genotype *CYP2D6*1/*1* is shown for all exons 1-9. The highest signal at each exon represented the junction probe (yellow bars). All other signals within exon 1-9 are less than 20% and have been neglected. That means that the main transcript in *CYP2D6*1/*1* is the conservative mRNA transcript including all exons (figure 17, A).

In contrast as shown before for genotype **41/*41* the jump-junction probe spanning from exon 5 to 7 showed a signal which represents about 40% of the whole mRNA transcripts. Additionally in **41/*41* intron retention was monitored, especially of whole or parts of intron 5 (30%), intron 6 (20%) as well as intron 1 (40%) (figure 17, B). Heterozygote genotype *CYP2D6*41/*1* showed a splicing pattern being mean of what was observed in *CYP2D6*41/*41* and *CYP2D6*1/*1* (not shown). The splicing patterns for haplotype *CYP2D6*2/*2* looked similar to *CYP2D6*1/*1* (compare figure 17, A and C). In case of *CYP2D6*4/*4* only one reverse transcribed RNA sample showed sufficient amounts of labelled cRNA. Hybridizing experiment of this sample

**4/*4* showed conservative splicing from exon 1-4. From there increased aberrant splicing was monitored between all other exons leading to intron retention (40-50%). That might be explained due to the mutation *1846G>A* positioned directly one bp before exon 4 (figure 17, D).

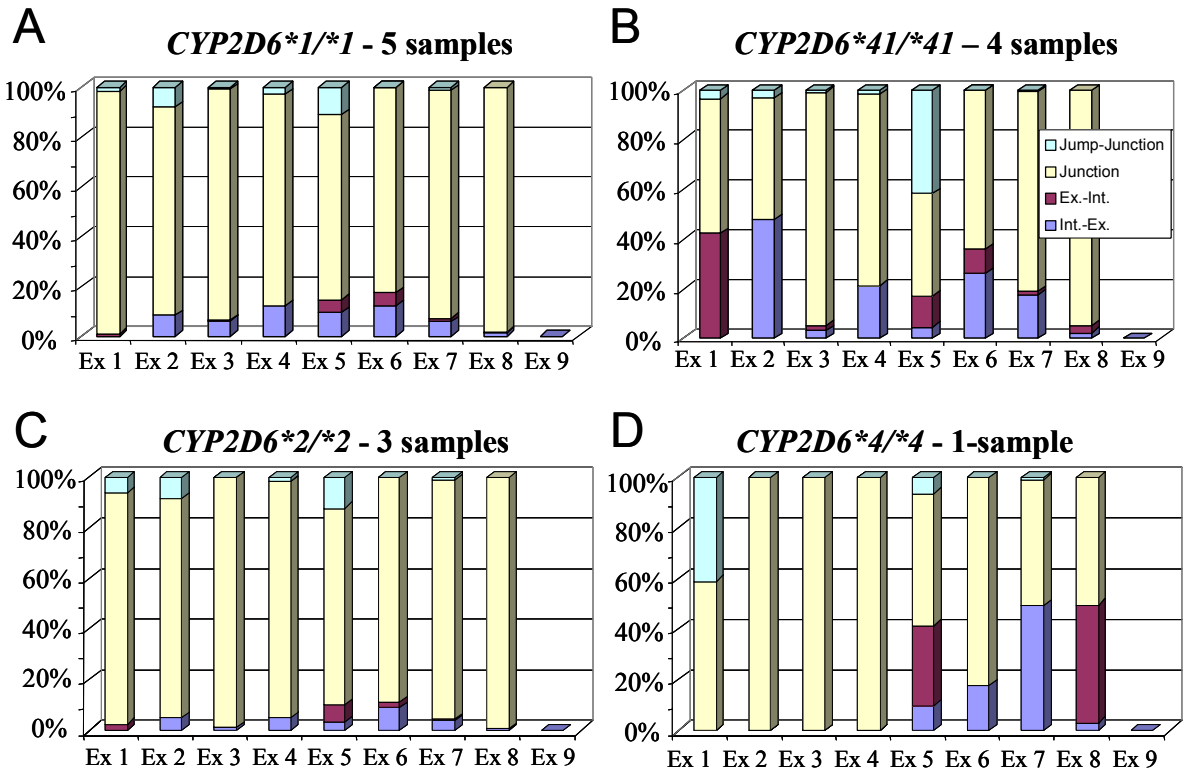


Figure 17: Hybridization experiments with target cRNA originating from 13 human liver samples with different genotypes.

Percentage of aberrant splicing transcripts $I_{fraction/100\%}$ shown after hybridization of 5 µg of target cRNA (experiment performed in duplicate): A: five human liver samples with genotype *CYP2D6*1/*1*; B: four human liver samples with genotype *CYP2D6*41/*41*; C: three human liver samples with genotype sample *CYP2D6*2/*2*; D: one human liver samples with genotype *4/*4.

Gene and pseudogene specific splicing variants were analysed (figure 18). In figure 18 A, C and E the results of hybridized cRNA originate from *CYP2D7* and 18 B, D, and F cRNA origin from *CYP2D6* is shown. In the first line the mean hybridization result origin from 8 liver samples is shown for *CYP2D7* (A) and for 10 liver samples for *CYP2D6* (B). The genotypes reversely transcribed for *CYP2D6* and hybridized were 2x**41/*41*; 1x**2/*2*; 2x**41/*1* and 5x**1/*1*. Samples used for pseudogene *CYP2D7* specific reverse transcription were the same as used for *CYP2D6*. (Two

samples carrying genotype *CYP2D6*1/*1* produced too less cRNA and weak Cy3 incorporation rates higher than 1000 NT/F (2.4.4.1). The two samples were not hybridized.)

Looking at the result seen in figure 18 A and B where the mean hybridization result of all 8 (*CYP2D7*) or 10 (*CYP2D6*) samples are shown, different splicing patterns can be seen comparing gene and pseudogene. Compared to *CYP2D6*, the pseudogene *CYP2D7* showed higher signals for splicing transcripts including intron 6 and a slightly higher signal for retaining of intron 4. In *CYP2D6* the jump-junction probe from exon 5 to 7 showed higher signal due to genotype **41,* which was included within the analysed pool of genotypes.

In the two lower panels of figure 18 C, D, E and F two liver samples (one with genotype **41* and one with genotype **1* are compared as single samples to demonstrate clear difference in splicing patterns also visible in single samples reverse transcribed for *CYP2D6* (figure 18 D and F) and for the pseudogene *CYP2D7* (figure 18 E and F). The results reflect the situation already observed within the pooled analysis. Interestingly the splicing variant skipping exon 6 (described for *CYP2D6*) was not detected in the 8 samples reverse transcribed for *CYP2D7.* Additionally the signals for retention of intron 1 which was increased within *CYP2D6*41/*41* samples was not seen in this height in *CYP2D7*. The experiments demonstrated that gene and pseudogene specific splicing exists, was detectable and can be distinguished with the developed splicing platform.

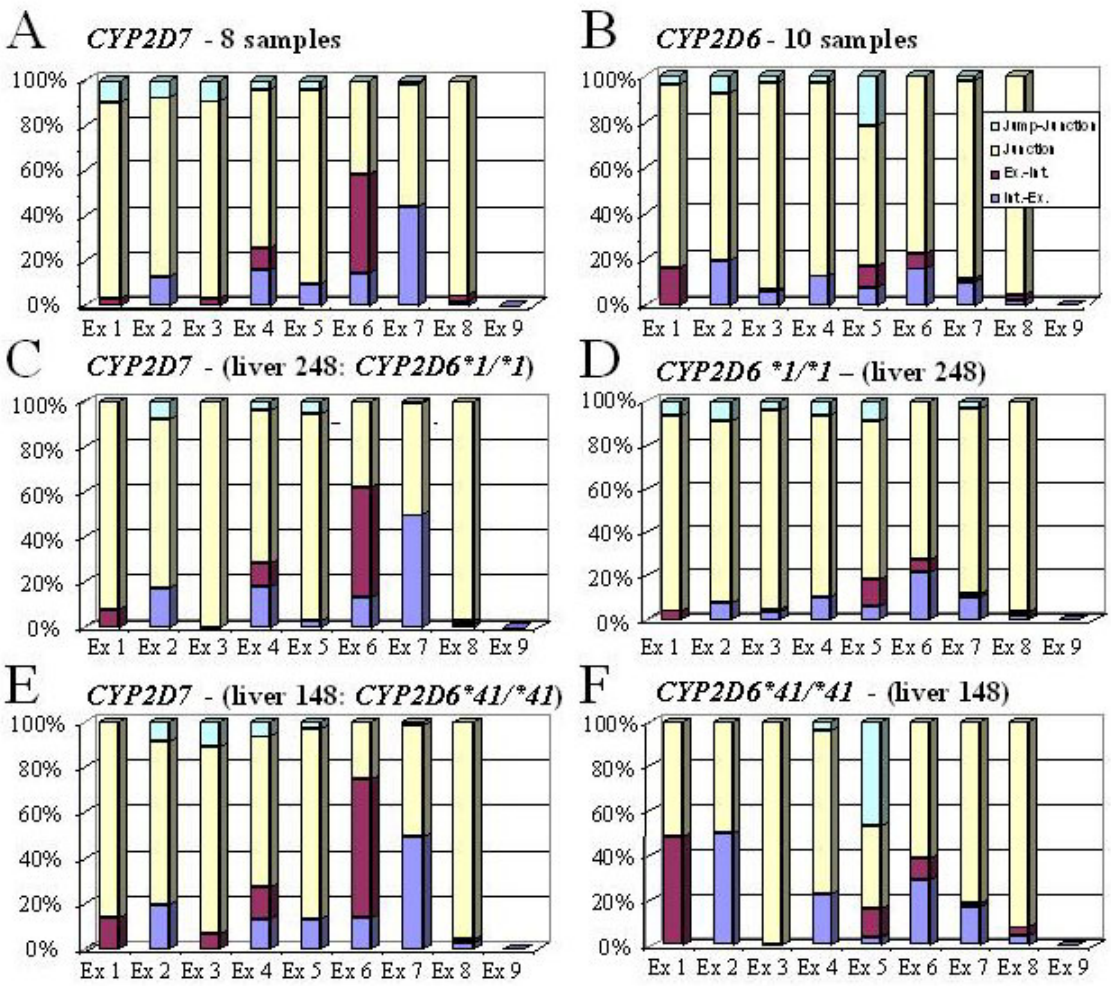


Figure 18: Hybridization experiments with target cRNA origin liver samples and reverse transcribed either for *CYP2D6* or *CYP2D7*.

Mean percentage of aberrant splicing transcripts $I_{fraction/100\%}$ shown after hybridization of 5 µg of target cRNA. A,C,E reverse transcription performed specific for *CYP2D7* and B,D,F specific for *CYP2D6*. (A+B): calculation of transcript diversity after individual hybridization in duplicate either *CYP2D7* (8 samples – A) or *CYP2D6* (10 samples - B). C, D: transcript diversity after hybridization of one liver sample with genotype **1/*1*. E, F: transcript diversity after hybridization of one liver sample with genotype **41/*41*.

3.2.2 Confirmation of microarray results using RT-PCR and sequencing

To verify the microarray results RT-PCR was used as an alternative method. Resulting PCR-fragments higher or lower in size as the expected normal fragment were cut, cloned and sequenced.

Two fragments of *CYP2D6* representing region exon 1 to exon 3/4 and exon 4/5 to exon 7/8 were amplified. For sample #148 (genotype **41/*41*) one fragment was present after amplification of region exon 1 to exon 4/3 and three fragments for sample #035 carrying genotype **1/*1*. In both samples the conservative fragment (with the expected size) was sequenced and showed inclusion of all exons from exon 1 to exon 3/4. A second fragment (larger than the conservative fragment) within genotype *CYP2D6*41* was not origin of *CYP2D6* or *CYP2D7* but NM_000818. No sequenced fragments included the intron 1, even though the splicing microarray showed increased amount of this variant in genotype *CYP2D6*41/*41*. Intron 1 retention is described in literature but not associated to any genotype (Gaedigk et al., 2005).

Within genotype *CYP2D6*1/*1* (sample #35) in addition to the conservative sequence a fragment lacking exon 3 was sequenced.

Analysing the RT-PCR fragments spanning exon 4/5 to exon 7/8, within genotype **41* two fragments were visible, the conserved full fragment including all exons in addition to one which had been shown to lack exon 6 by sequencing.

Within genotype *CYP2D7* three fragments with intron retention were defined one including 194 bp of 5'intron 6, one including 57 bp of 3'intron 6 and one including the whole intron 6 region. This result is in agreement with the hybridization result of *CYP2D7*, where higher amounts of intron 6 were monitored. Putting this together the sequencing reflect the microarray result although not all splice variants defined in the literature and expected from the hybridization results have been confirmed by sequencing.

3.3 Aberrant splicing in *CYP2B6*

Cytochrome *2B6* is highly polymorphic with 28 alleles and over 100 SNPs described in the literature. *CYP2B6* expression (mRNA, protein and activity) is interindividually highly variable and the mechanisms behind this are not totally clarified. A number of indications for splice variants of *CYP2B6* are described in the literature as shown in figure 5 but compared to *CYP2D6* no genotype dependent splicing was described so far. To investigate alternative splicing within *CYP2B6* in more detail the developed splicing platform was transferred to *CYP2B6*.

3.3.1 Transfer of splicing microarray platform to *CYP2B6*

CYP2B6 gene specific probes were designed using the same borderlines used for *CYP2D6* (Tm, GC content and length) (chapter 3.1).
After spotting the microarray evaluation was performed with amplified and labelled target DNA including either only exons (amplification of previously published *CYP2B6* cDNA vector: pIKAT15, (Lang et al., 2001) or intronic regions (amplification of genomic DNA originate from a liver sample previously genotyped as **1/*1*). Two fragments including only exonic regions were successfully amplified, the first spanning exon 1 to exon 6/7 (primers: cDNA2B6FE2 and cDNA2B6RE7) and the second including exon 5 to exon 9 (primers: cDNACYP2B6FE5 and cDNACYP2B6RE9). Additionally six short fragments including intronic regions were amplified labelled and hybridized to evaluate the intron-exon probes: exon1 to intron 1 (primers: CYP2B6-1F and CYP2B6-1R), intron 1 to intron 3 (CYP2B6-2F and CYP2B6R), intron 3 to intron 4 (CYP2B6-4F and CYP2B64R), intron 4 to intron 5 (CYP2B6-5F and CYP2B6-5R), intron 5 to intron 8 (CYP2B6-5F and CYP2B6-8R), intron 8 to exon 9 (CYP2B6-9F-CYP2B6-9R). The same hybridization conditions were used as for *CYP2D6* and they were shown to be optimal. Five probes were exchanged as they showed unspecific hybridization results. In those five cases higher hybridization specificity was observed using a 3'oligo-dT-C7-Aminolink (marked with C7-aminolink in appendix/ table 1). That might be due to sterical changes during hybridization by placing the oligo dT-linker either on the 3' or 5'prime site of the probe. As next step RNA from liver samples was reverse transcribed and linear amplified using the *CYP2B6* gene specific T7 primer (*CYP2B6*-T7). This T7-

primer shows sequence differences to *CYP2B7*. The reverse transcription of *CYP2B6* led to reduced amounts of cRNA but similar labelling efficiency (NT/F) compared to *CYP2D6*. After hybridization of *CYP2B6* specific cRNA the background aberrations were slightly higher and the signals of the *CYP2B6* probes were lower compared to *CYP2D6*. Several repetitions of linear amplification and hybridizations were necessary to achieve results, which could be analysed. In figure 19 the hybridization result of cRNA origin from genotype **1/*1* and from a sample carrying **6/*6* were reverse transcribed and hybridized. It can be seen that in genotype *CYP2B6*1/*1* the conservative full functional mRNA including all exons is the most abundant form (figure 19, A). For *CYP2B6*1/*1* not much additional splicing can be observed, at least none which was present higher than 20%. In contrast to *CYP2B6*1/*1* for CYP2B6**6/*6* (figure 19, B) it can be seen that the jump-junction probe from exon 3 to 5 showed an increased signal (>20%). This indicates that exon 4 was spliced out. The signal of the probe for intron to exon 4 was also strongly increased from nearly 20 to 40%.

No increased signal was observed for the probe exon 3 to intron 3, the result for the probe intron-exon4 was interpreted as partial retention of intron 3. Furthermore the probe exon 8 to intron 9 is slightly increased assuming that intron 8 might be retained in some transcripts.

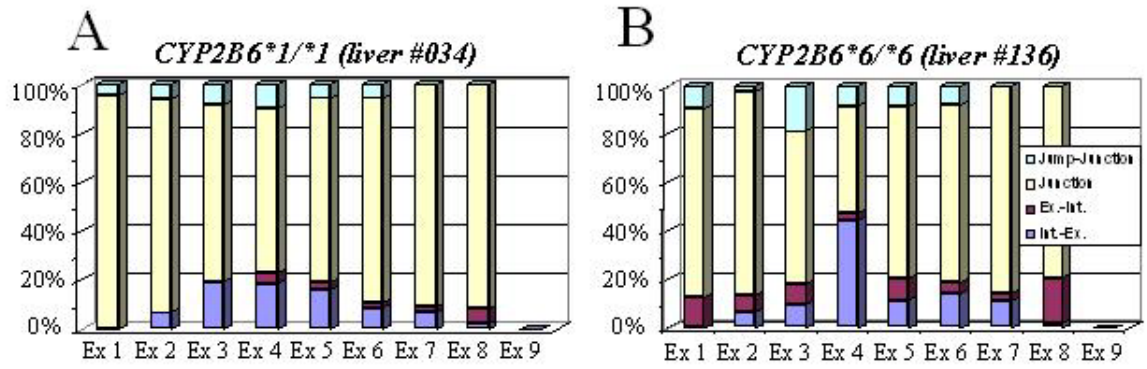


Figure 19: Hybridization experiments with target cRNA reverse transcribed for *CYP2B6* from human liver samples. Percentage of aberrant splicing transcripts $I_{fraction/100\%}$ shown after hybridization of 5 µg of target cRNA (in duplicate) originating from A: one human liver sample with genotype *CYP2B6*1/*1*; B: one human liver sample with genotype *CYP2B6 *6/*6*

3.3.2 Alternative splicing of *CYB2B6* pre-mRNA in human liver

In order to confirm splicing results within *CYP2B6* obtained with the microarray two RT-PCR assays were designed: one amplifying a region between exon 2/3 and exon 7 and a second amplifying a region between exon 5/6 and exon 9. The region upstream of exon 2 was not considered because no splice variants have so far been described in this region or seen with the *CYP2B6* splicing array (figure 19) and because of experimental obstacles due to the size of intron 1 (12.7 kb).

RT-PCR between exon 2/3 and exon 7 resulted four fragments. The largest fragment, the normal product (NP) (figure 20, A1 and A2) was identified by sequencing as the mRNA transcript that corresponds to the reference cDNA sequence. Two additional fragments seen in this RT-PCR were identified as novel splice variants (SV). SV7 lacked exon 4 and exon 5 except for a residual 50 bp fragment whereas SV8 lacked only exon 4 (table 2, 20 A1 and A2).

Additional quantification of specific RT-PCR transcripts on the Agilent Bioanalyzer showed significant higher NP in samples with *CYP2B6*1/*1* with reduction towards **1/*6* ($p<0.05$) and **6/*6* ($p<0.05$). SV1 was hardly detectable in *CYP2B6*1/*1* but visible in **1/*6* and **6/*6*. Within samples with *CYP2B6*1/*6* the SV1 and NP showed comparable amounts (figure 20, B).

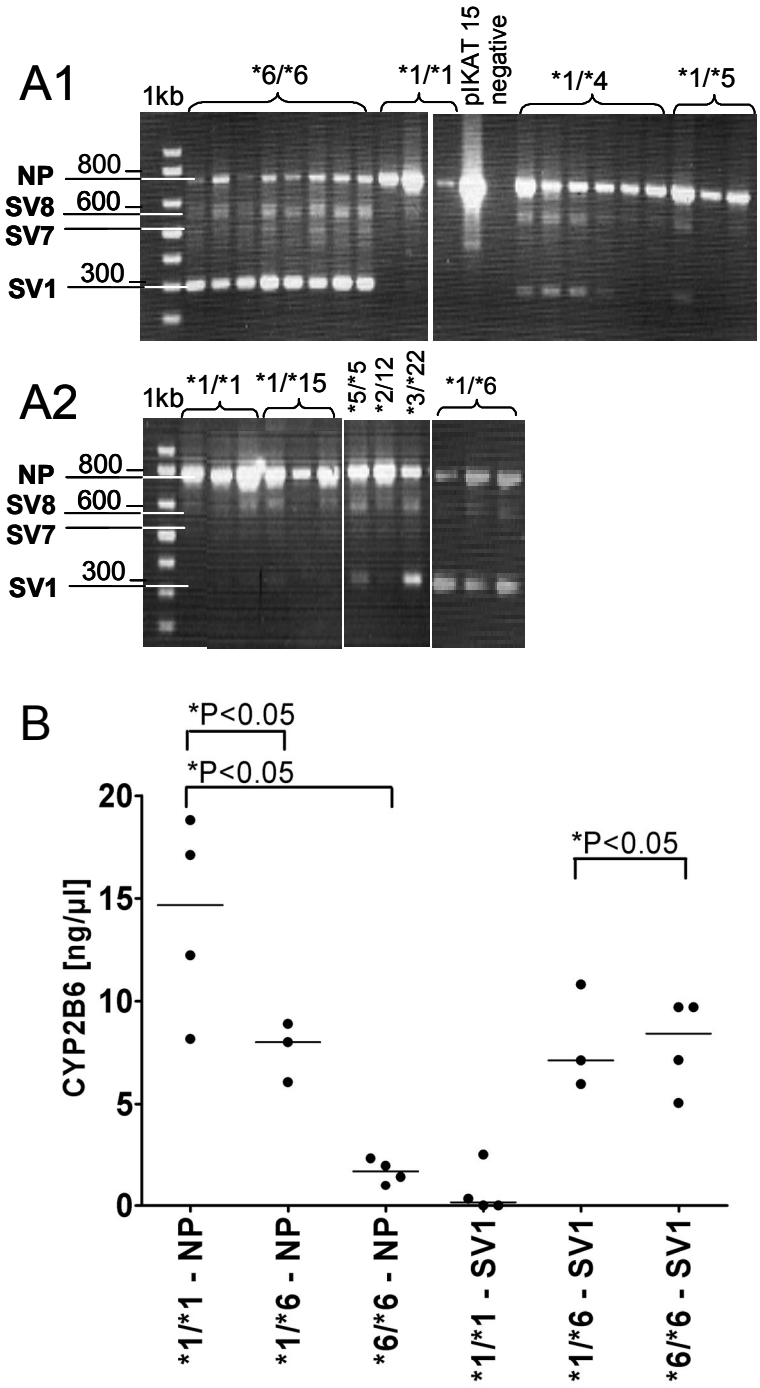


Figure 20: RT-PCR Analysis of *CYP2B6* aberrant splicing in human liver tissue

A1 and A2: Agarose gel of RT-PCR product of liver samples previously genotyped for *CYP2B6*. Positive control: cDNA vector pIKAT15; negative control (no DNA); amplified regions: exon 2/3 to exon 7. NP, normal product, representing the full-length functional transcript; SV1, SV8 refers to splice variants identified by sequencing (see table 1).

B: Quantification of specific transcripts. The purified RT-PCR products (see A) were analysed on the Agilent Bioanalyzer. The amounts of complete fragment NP including all

exons from 2-7 and splice variant SV1 lacking exon 4-6 were determined relative to an internal standard added to each sample before running the DNA lab-on-a chip assay.

The presence of SV8 was indicated by the *CYP2B6* splicing microarray (figure 19, B) by the higher signal intensities of the jump-junction probe for exon 3 to 5 comparing *CYP2B6*1/*1* with *CYP2B6*6/*6*. Concerning this fragment with the RT-PCR analysis, no clear genotype dependent increase was detectable for this fragment. The smallest fragment (284 bp) was confirmed to be identical to the variant SV1 lacking exons 4 to 6. Whereas SV7 and SV8 were apparently not genotype dependent, the intensity of NP and SV1 varied strongly according to genotype. SV1 was the most abundant fragment in samples carrying genotype **6/*6*, but was hardly detectable in **1/*1* homozygotes. Heterozygous samples (**1/*6)* had almost equal amounts of NP and SV1. The amount of NP was highest in **1/*1*, intermediate in **1/*6* and lowest in **6/*6* genotypes (figure 20). Interestingly, several samples with genotype **1/*4* and a single sample with genotype **3/*22* also had clearly detectable amounts of SV1 (figure 20, A1 and 20, A2).
The RT-PCR clearly showed the presence of SV1 and SV7. These splice variants were not detected with the microarray. This is due to the use of jump-junction probes, which display only the skipping of one single exon. The skipping of multiple exons as present in SV1 and SV7 is therefore not displayed with the here used microarray design.

RT-PCR analysis of the RNA region spanning exon 5/6 to exon 9 resulted in three transcripts present in all genotypes (figure 21). Sequencing of these fragments revealed besides the normal product, two variants: one lacking exon 8 but included an alternative exon 8A (λMP8, figure 21) and a second transcript, so far not described, which lacked exon 8 (SV9, table 2). These two variants were apparently not genotype dependent and where also not monitored with the *CYP2B6* microarray. As by-product during sequencing it was found that in one sample (liver #251) with genotype **1/*1* the mutation 18273G>A within intron 5 was found, which was so far only once reported within the Japanese population (Klein et al., 2005).

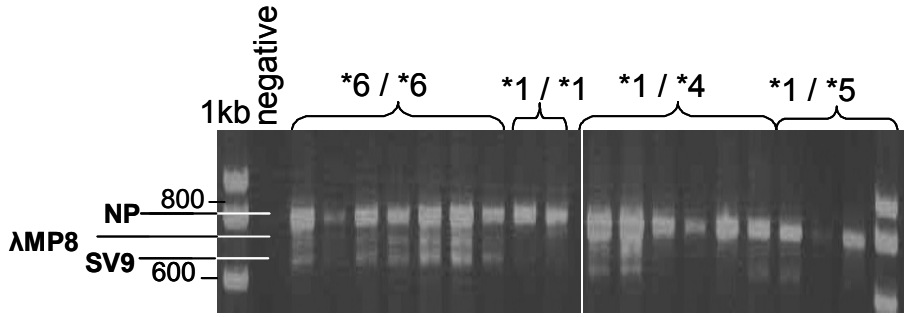


Figure 21: RT-PCR analysis of *CYP2B6* aberrant splicing in human liver tissue
Regions between exon 5/6 and exon 9 were analysed. NP, normal product, representing the full-length functional transcript; λMP8 and SV9 refer to splice variants identified by sequencing (see table 2 and 4).

Table 2: New splicing variants within *CYP2B6*
Shaded boxes represent exonic sequences. SV7, SV8 and SV9 are novel splice variants not described before. Sequence numbers origin from sequence NG_000008.5 with ATG start codon set as 1.

name	mRNA transcript	description
SV7	2 | 3 | 50 bp of exon 5 | 6 | 7	lack of exon 4 only 50bp of exon 5:17914-17963
SV8	2 | 3 | 5 | 6 | 7	lack of exon 4
SV9	5 | 6 | 7 | 9	lack of exon 8

3.3.3 Investigation of mRNA variability using a TaqMan assay

With the knowledge, achieved by RT-PCR analysis that the most common alternative splicing event within *CYP2B6*6* shows skipping of exon 4-6 the mRNA quantification achieved with a previously developed TaqMan assay (Zukunft et al., 2005) was analysed concerning alternative splicing. The specificity of this assay was controlled with pCR4-vectors in which the purified alternatively spliced fragments achieved through RT-PCR analysis were cloned and sequenced (table 3). The assay showed specificity for all mRNA transcripts, which include exons 4 and 5 (pIKAT15 and pMHH-CYP2B6-TOPO-134) and no signal for transcripts lacking exon 4 or exon 4-6

(pMHH-CYP2B6-TOPO-98, -TOPO-101, -TOPO-111, -TOPO-207 and -TOPO-215) (table 3).

Table 3: Plasmids including *CYP2B6* splicing transcripts

Vector	**Properties**	**Origin/ Reference**
pIKAT15	*CYP2B6*-cDNA vector	Lang et. al., 2004
pMHH-CYP2B6-TOPO-98	pCR4-vector containing splice variant SV1 of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-101	pCR4-vector containing splice variant SV8 of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-134	pCR4-vector containing conservative splice product of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-111	pCR4-vector containing splice variant SV7 of *CYP2B6* between exon 3-7	this study
pMHH-CYP2B6-TOPO-207	pCR4-vector containing splice variant SV9 of *CYP2B6* between exon 5-9	this study
pMHH-CYP2B6-TOPO-215	pCR4-vector containing splice variant λMP8 of *CYP2B6* between exon 5-9	this study

That means quantification of mRNA from 150 human liver samples only detects full functional reference splice product including exon 4 and 5.

CYP2B6 mRNA expression was significantly lower in samples with **1/*6* ($p<0.05$, n=42) or **6/*6* ($p<0.01$, n=6) genotype compared to **1/*1*. So this pattern reflects the situation observed in RT-PCR. Lower amount of mRNA expression within **1/*6* and **6/*6* might be explained by the increased amount of alternative splicing transcript SV1 and deduced amounts of NP as seen in analysis of RT-PCR within those samples. All other investigated genotypes with n>3 (**1/*5, *5/*6, *4/*5, *1/*2*) showed no significant difference on mRNA level compared to **1/*1* (figure 22). This fits well with the RT-PCR results (figure 20, A1 and A2)

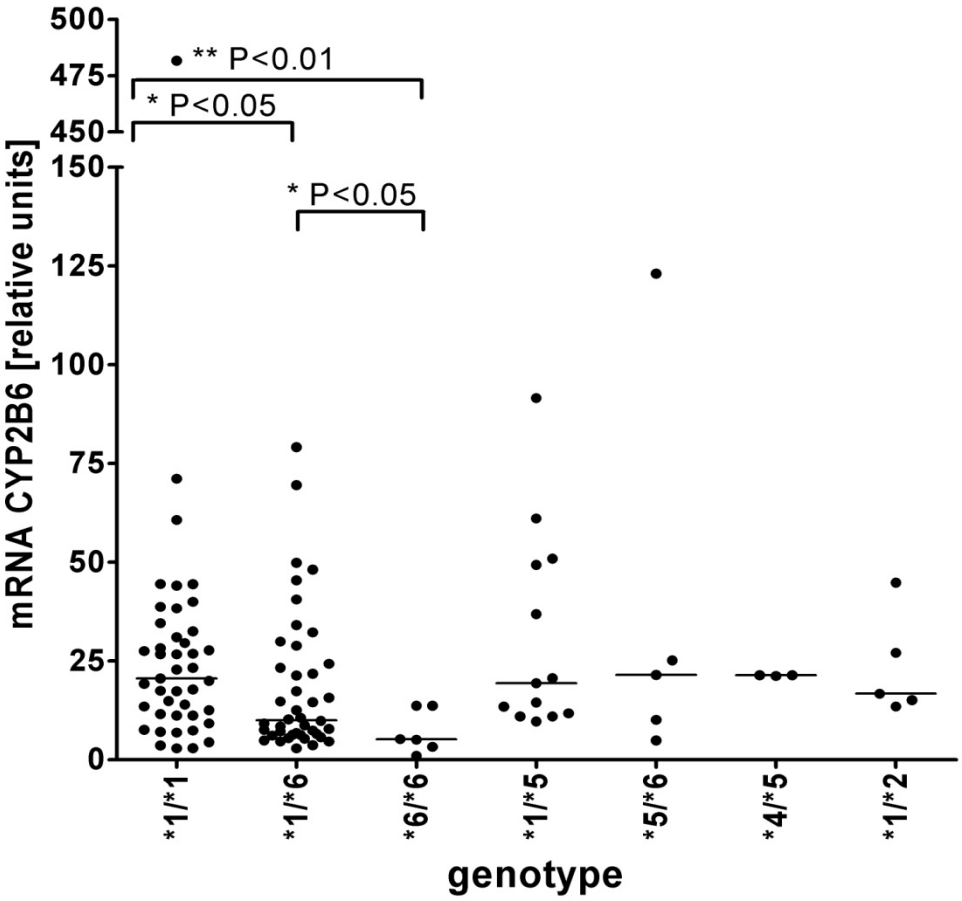


Figure 22: Relationship between genotype and *CYP2B6* transcript levels in human liver samples. A: Scatterplots depicting β-actin normalized *CYP2B6* mRNA determined by real-time RT-PCR for various genotypes. The lowest value was arbitrarily set at 1, i.e. values represent fold differences compared to the lowest expression.

3.3.4 Influence of intron mutation on aberrant splicing

As it was described in literature intron 3 15582C>T had been suggested to be potentially involved in erroneous splicing (Lamba et al., 2003). Therefore all 235 human liver samples were genotyped for this mutation using MALDI-TOF MS technology. Haplotype analysis showed that 15582C>T was not present in any of the homozygous **6/*6* samples (n=16) but associated with alleles **1C*, **13B* and **15*. The levels of mRNA, protein expression and enzyme activity were not significantly different but similar in wild type samples with or without this intronic mutation (shown for mRNA in figure 23).

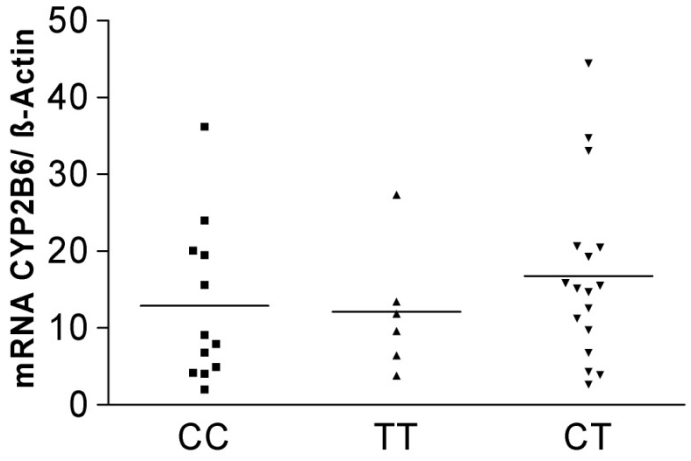


Figure 23: Influence of SNP 15582C>T on *CYP2B6* transcript levels in human liver samples (with *CYP2B6*1/*1* background).

In RT-PCR assay no difference in NP or SV1 was observed when **1/*1* 15582CC was compared to **1/*1* 15582TT (figure 24). In samples genotyped **1/*1* low amounts of SV1 was detected but normal amounts of NP (figure 24).

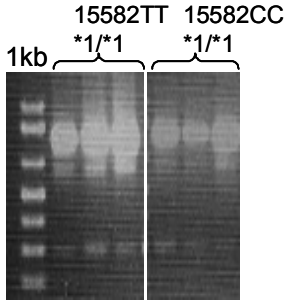


Figure 24: Influence of SNP 15582C>T analysed with RT-PCR
Analysis of a fragment including exon 2/3 to exon 7. All samples shown have a *1/*1 background.

3.4 A minigene approach to investigate *CYP2B6* splicing

To unequivocally identify the mutation(s) that are causally responsible for the decreased expression phenotype of the **6* allele a series of minigene constructs (figure 25 to 27) were designed. The aim was to create constructs, which include the entire open reading frame so that effects of mutations could be analyzed on the mRNA, protein and activity level. Due to the large size of some of the introns the cloning strategy included deletion of introns 1, 4 and 8 and parts of introns 3 and 6.
Two overlapping fragments, the first comprising exons 2 to 4 (3073 bp), and fragment 2 comprising exons 5 to 8 (3877 bp), were amplified from *CYP2B6*1/*1* genomic DNA samples (figure 25, A). An internal fragment of 1437 bp was excised from intron 3 using the restriction enzyme *Mfe*I. Intron 6 was also shortened by a fragment of 1535 bp using enzyme *Bsg*I. Both fragments 1 and 2 were digested with *Eco*NI and *Sac*I restriction enzymes and were then combined with *Eco*NI-digested pIKAT15 (figure 25, B) in a three fragments ligation reaction (figure 25, C). This resulted in the parental minigene construct [*1]. A previously described *CYP2B6* cDNA vector (pIKAT15) was digested with *Hind*III and *Sbf*I and the insert was replaced by a 518 bp fragment (exons are shaded in dark grey) to include an additional *Eco*NI restriction site by the reverse primer (figure 25, C-D).

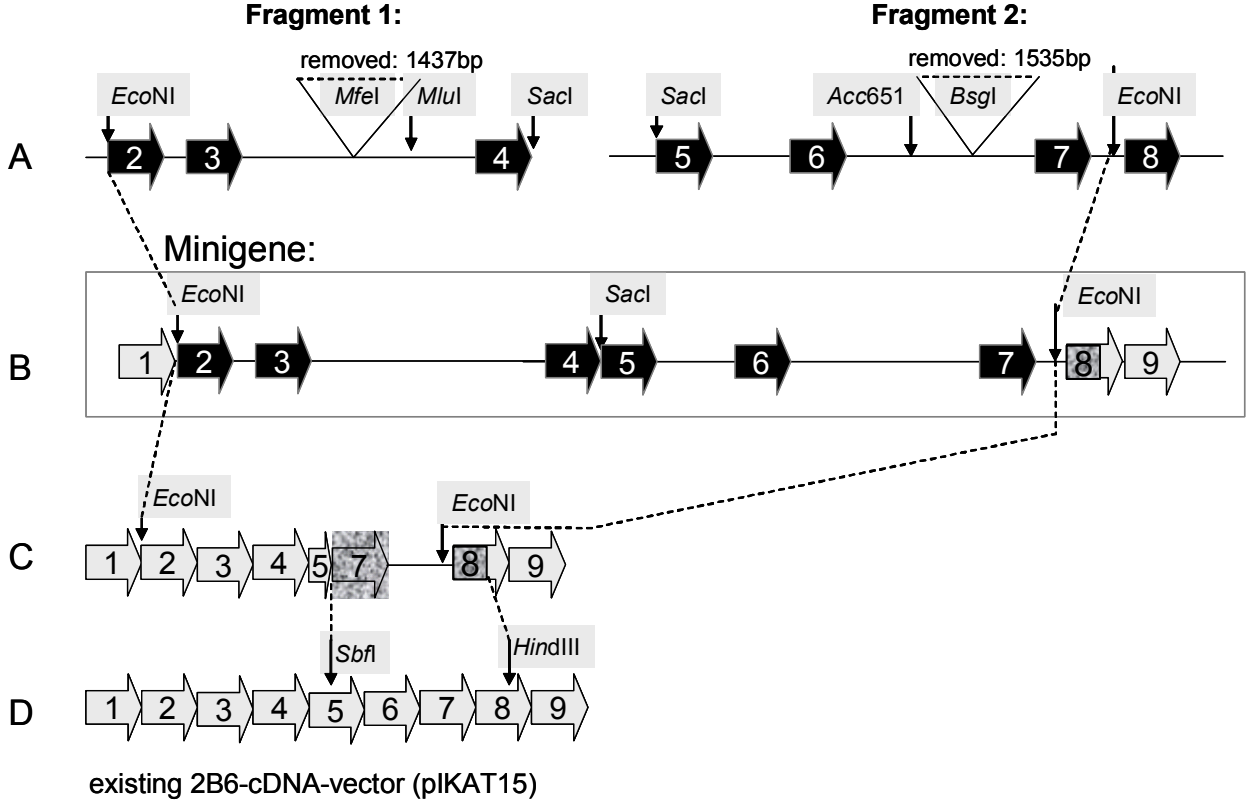


Figure 25: Construction of the *CYP2B6* minigene
Exons of amplified fragments are shown in black. Exons remaining from pIKAT15 are shown in light grey.

The final constructs thus contained all 9 exons as well as the 5'- and 3'-parts of introns 3 and 6 and the complete introns 2, 5, and 7 and were cloned in pCMV4-vector (figure 26).

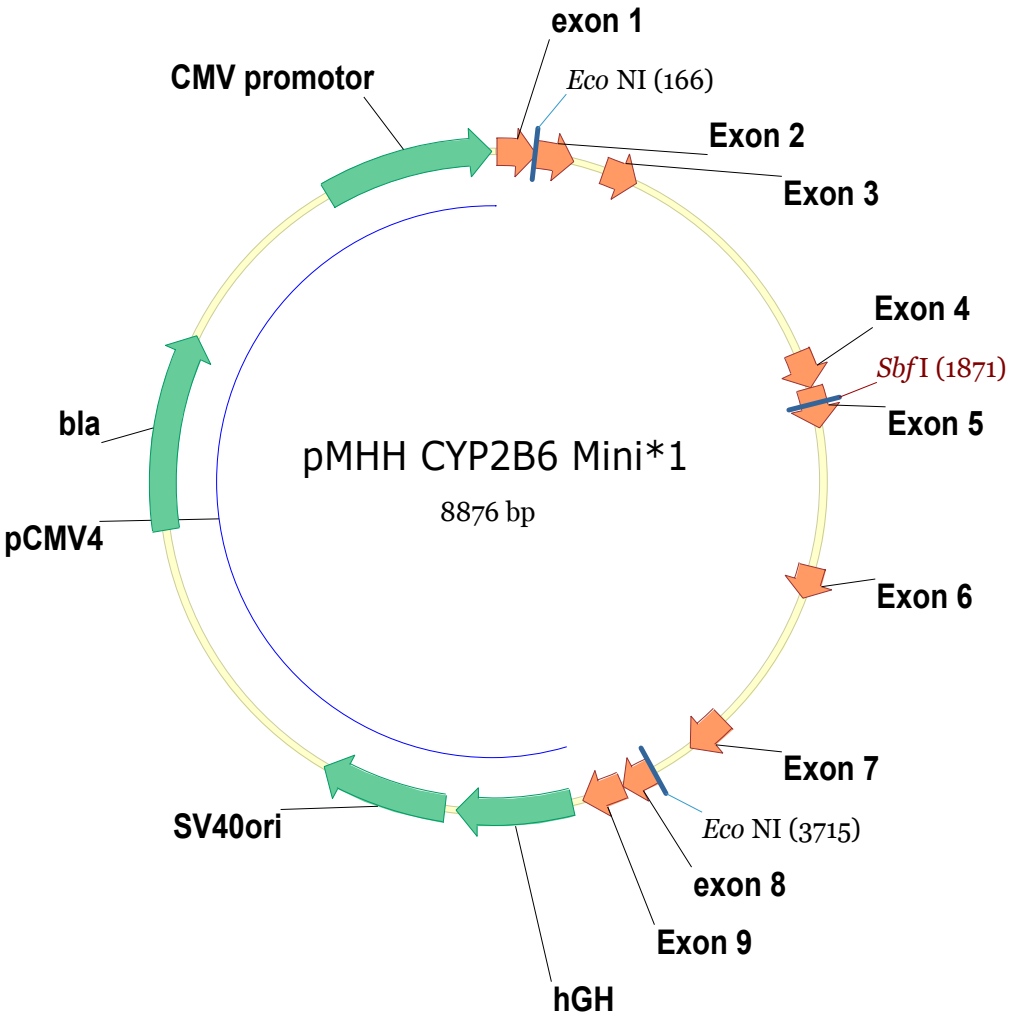


Figure 26: Structure of the *CYP2B6* minigene in the mammalian expression vector pCMV4

In vitro mutagenesis was used to introduce relevant SNPs (figure 27). To elucidate the causative mutation in the **6* allele, both SNPs c.516G>T and c.785A>G were introduced individually (minigenes [*9] and [*4], respectively) as well as combined (minigene [*6]). In addition a minigene representing the so far functionally uncharacterized c.777C>A [*3] was constructed, which also appeared to be associated with SV1 (see above, figure 20, A2).

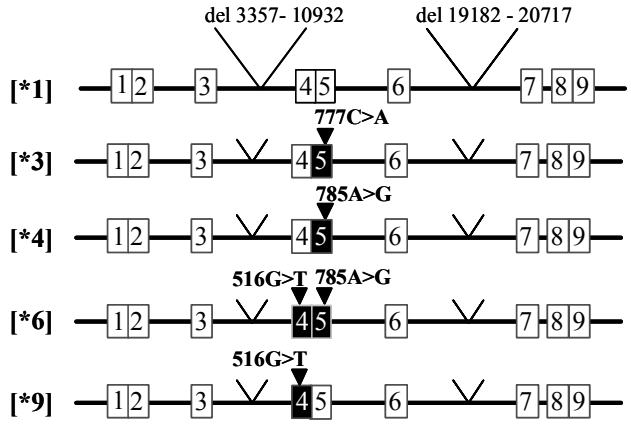


Figure 27: Structure of *CYP2B6* minigene constructs. All constructed minigenes are shown with their mutations indicated

3.4.1 Establish transfection experiments for studying alternative splicing

Initial transfection experiments with Huh7, a human hepatoma cell line, and COS-1, an African green monkey kidney epithelial cell line, resulted in different amounts of the normal product (NP) between the various constructs. However SV1 was generally produced at high amounts, almost no NP was seen and no differences were seen between the minigene constructs after transfection of 1.5 µg/ $2.5.10^5$ cells (Huh7-1.5 µg/ figure 28). In addition, other splice variants also appeared to be produced at much higher amounts compared to liver. This is probably due to an overflow of the splicing machinery when too much DNA is transfected into the cells.
However, when the plasmid DNA was titrated into empty vector, gradual changes of the relative amounts of NP and SV1 were observed. The lowest amount of transfected minigene (0.0015 µg), resulted in almost no SV1 for [*1] and also less other splice products (Huh7-0.0015 µg/ figure 28).

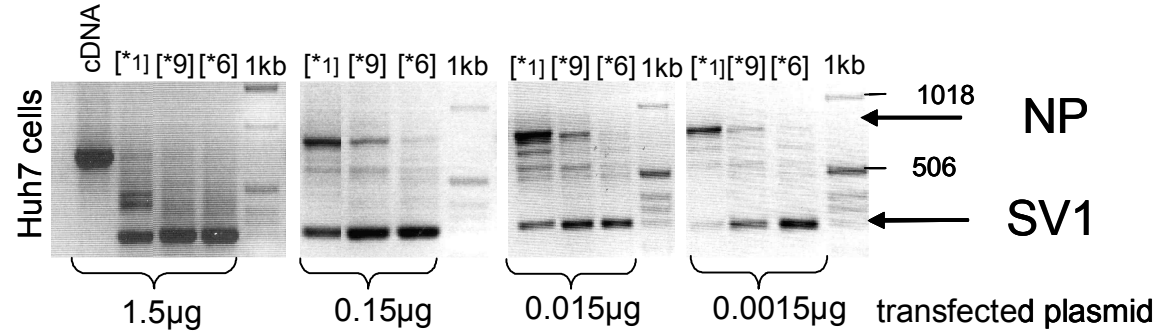


Figure 28: Transfection experiments with Huh7 cells using different amount of transfected plasmid. 24 h after transfection with minigenes total RNA from Huh7 cells transfected with minigene constructs was reverse transcribed and PCR amplified with primers located in exon 2/3 and exon 7. Transfection experiment performed with minigenes [*1], minigene [*9], minigene [*6] as well as 1.5 µg of cDNA vector pIKAT15.

3.4.2 Analysis of alternative splicing in minigene-transfected cell lines

The mRNA transcripts formed in COS-1 and Huh7 cells, transfected with 1.5 ng of the minigene constructs were analyzed by the RT-PCR within the region exon 2/3 to exon 7. As shown in figure 29, the NP fragment was the most abundant in the [*1] transfected cells and only small but clearly detectable amounts of SV1 were produced. In contrast, the SV1 fragment was most intense in cells transfected with [*6] and [*9], which produced in contrast only low amounts of NP.

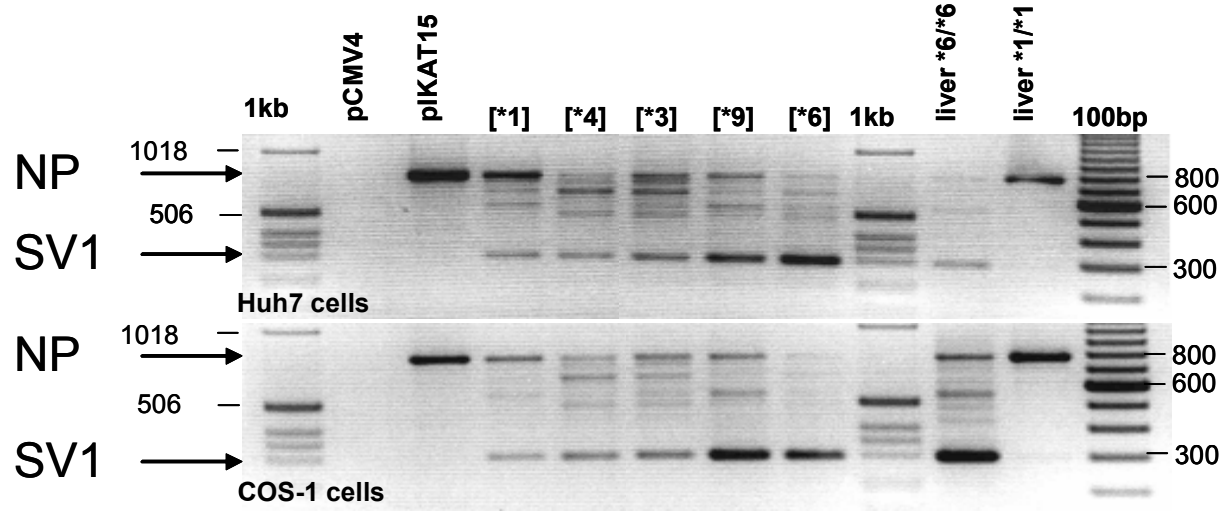


Figure 29: Total RNA from Huh7 and COS-1 cells transfected with 1.5 ng minigene constructs was reverse transcribed and amplified with primers located in exon 2/3 and exon 7.

Transfection of all five minigene constructs resulted in immunologically detectable amounts of full-length CYP2B6 protein, as shown by Western blot analysis (figure. 30, A1 and A2). Quantification of β-gal normalized CYP2B6 protein revealed approximately similar expression of [*1]- and [*4]-minigene in Huh7 and COS-1 cells. In contrast, both constructs harbouring the c.516G>T SNP, namely [*9] and [*6], consistently expressed very low amounts of protein. The presence of c.785A>G appeared to slightly increased expression (compare [*4] with [*1], and [*6] with [*9] in

figure. 30, A1 and A2). The [*3] minigene construct expressed about one third of CYP2B6 protein compared to the reference construct [*1].

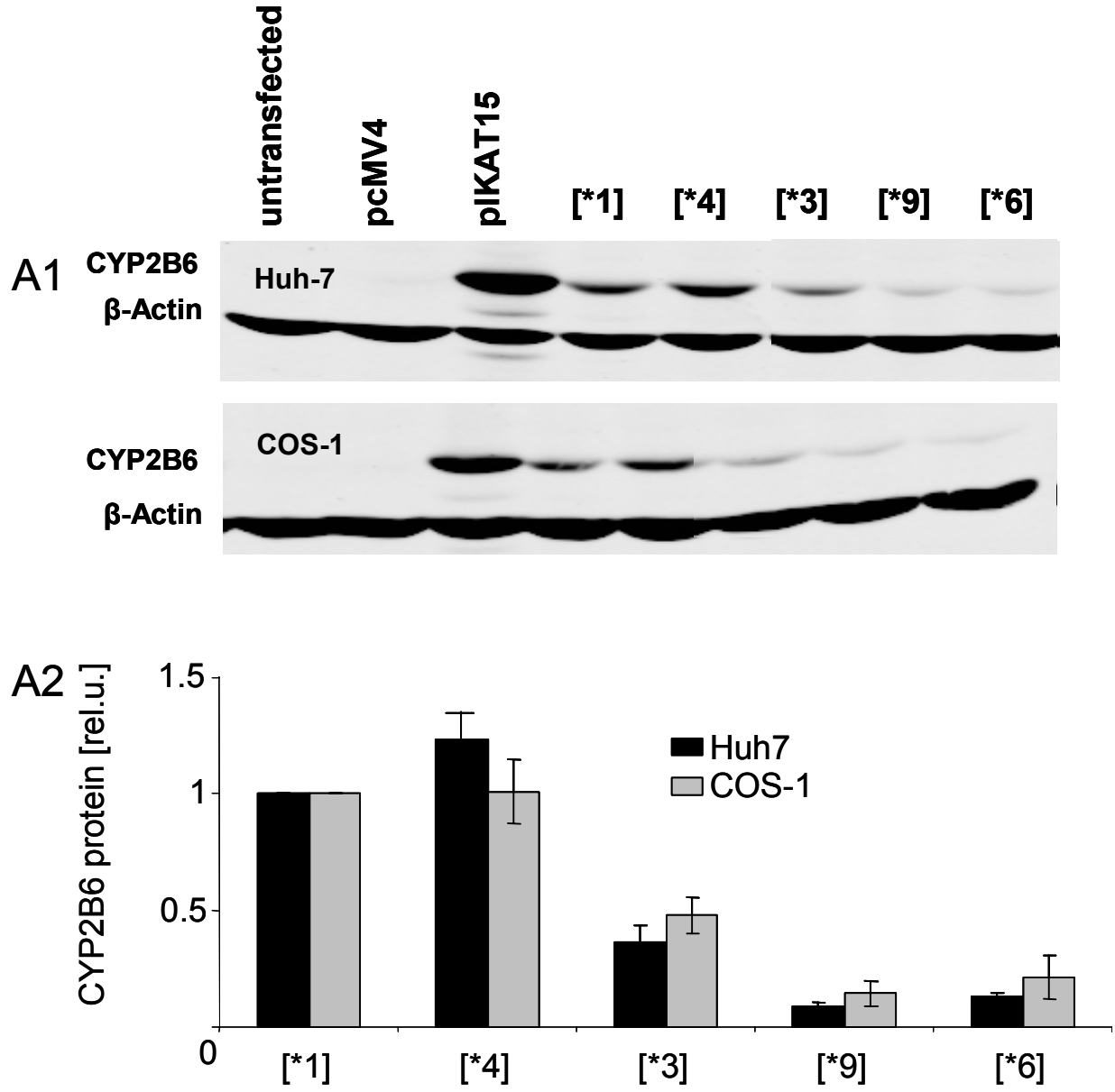


Figure 30: Western blot analysis of transfection experiment

A1: 48 h hours after transfection of minigenes into Huh7 and COS-1 cells CYP2B6 and β-actin were analyzed by immunoblotting using specific monoclonal antibodies. Transfection efficiency was additionally controlled by measuring β-galactosidase activity.

A2: CYP2B6 protein was quantified from the immunoblotting results of 3 independent transfections and normalized to β-galactosidase activity measured in cell extracts. The mean values with indicated standard deviations are shown relative to the mean protein amount for minigene [*1] set at 1.0.

Bupropion hydroxylase activity was only measurable in transfected COS-1 cells and closely followed the protein levels. Activities of transfected Huh7 cells were too low to be reliably determined. Notably, the activities of the [*6] and [*9] constructs were extremely low whereas the [*3] construct showed an intermediate activity level (figure 31). The activity of the [*4] construct was slightly higher compared to the [*1] construct.

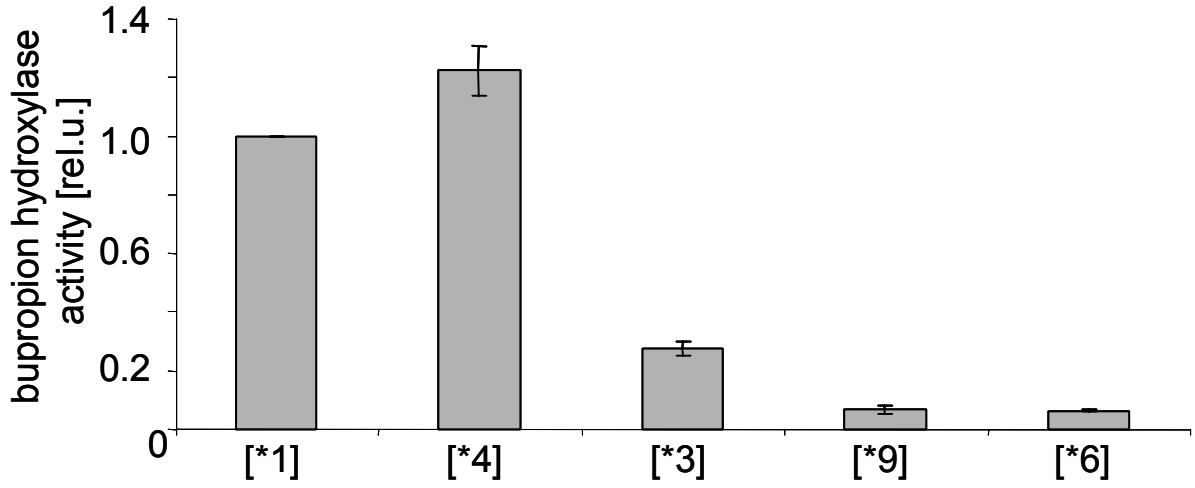


Figure 31: Bupropion hydroxylase activities

Activity was determined in protein extracts originating from COS-1 cells and normalized to β-galactosidase activity. Shown are the mean activities of doublicate measurements.

To summarize these data the mRNA results (figure 29) are consistent with the protein and activity data of the transfected cells (figure 30 and 31) and with the observations made in homozygous **1/*1* and **6/*6* human liver samples, although generally the extent of alternative splicing appeared to be significantly increased in transfected cells. Compared to the [*1]- and [*9]-constructs, the [*3]-construct resulted in intermediate amounts of NP and SV1 transcripts.

3.5 In silico analysis of alternative splicing within *CYP2D6*41* and *CYP2B6*6* using the ESE-finder algorithm

Using the ESE-finder algorithm (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi ?process=home) (Cartegni et al., 2003; Smith et al., 2006) a sequence window of 20 nucleotides surrounding the *CYP2D6* SNPs 2850C>T and 2988G>A were reanalysed for putative exonic splicing enhancer motifs using ESE-finder 3.0 as described by (Toscano et al., 2006; Zanger and Hofmann, 2008).

The SNP mutation within *CYP2D6*41* 2988G>A was shown to change the scores for binding of certain splice factors to known splicing enhancer elements, as illustrated in figure 32. The c.2988G>A SNP results in the loss of two binding sites for SF2/ASF (IgM-BRCA1) and a new predicted binding site for SRp40. The 2850C>T SNP (R296C of *2) in exon 6 leads to the loss of binding sites for SRp55 and for SF2/ASF whereas a new binding site for SC35 is predicted (Zanger and Hofmann, 2008).

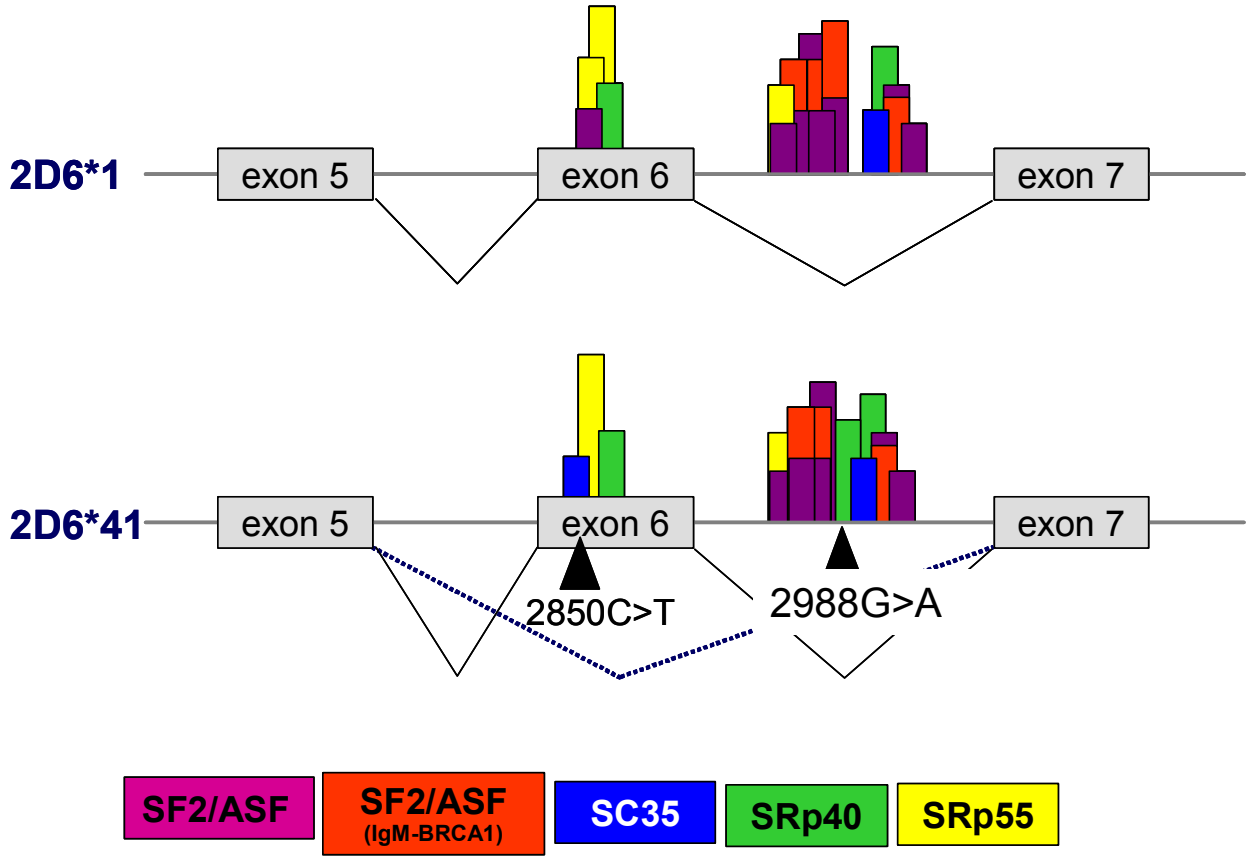


Figure 32: *In silic*o analysis of *CYP2D6* pre-mRNA splicing factor binding sites in exon 6 and intron 6. Predicted binding sites are shown schematically for the indicated alleles. The heights of the bars indicate the strength of binding sites as calculated by the ESE-finder. ESE, exonic splicing enhancer, SR, serine-arginine rich protein: SC35, SRp40, SRp55 and SF2/ASF.

The same method was used for *in silico* analysis of the 20 nucleotides surrounding the key mutation within *CYP2B6*6* to find bindings sites of splicing factors within exon 4 and 5 which might be influenced by the mutations. It was shown that c.516G>T SNP in exon 4 leads to the predicted loss of two binding sites for ESE-binding factors SC35 and SRp40. This probably results in the modulation of splicing events in favor of an alternative process that results in the observed splice variant lacking exons 4, 5 and 6. The c.785A>G SNP in exon 5 results in only minor changes of binding strength and the creation of one additional SF2/ASF (IgM-BRCA1) binding site (figure 33) (Zanger and Hofmann, 2008).

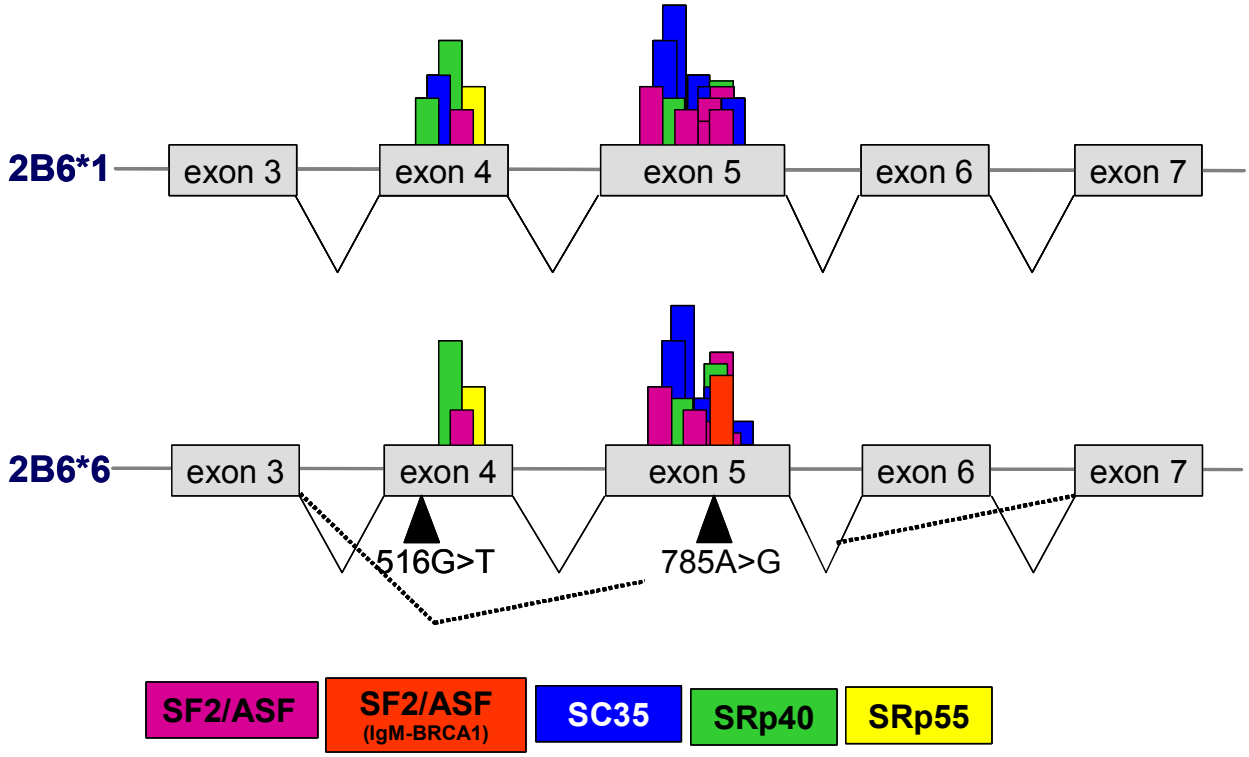


Figure 33: *In silico* analysis of *CYP2B6* pre-mRNA splicing factor binding sites in exons 4 and 5. Predicted binding sites are shown schematically for the indicated alleles. The height of the bars indicates the strength of binding sites as calculated by the ESE-finder. ESE, exonic splicing enhancer, SR, serine-arginine rich protein.

4 Discussion

This study is devoted to the detection of alternative splicing and transcript pattern within the polymorphic drug metabolizing cytochromes P450 CYP2B6 and CYP2D6. In the first part of this work a microarray sensitive for analysis of aberrant splicing was developed and used for detection of splicing patterns within *CYP2D6* in comparison to the pseudogene *CYP2D7*. Within *CYP2D6* the allele **41* explains the IM phenotype by the enhanced skipping of exon 6. This was used as an example to evaluate the developed splicing platform. In addition to the functional *CYP2D6* mRNA (including all exons), 13 variant transcripts are known for *CYP2D6* and *CYP2D7* (table 1). The microarray was used to clarify if allele specific, as well as gene and pseudogene specific transcript patterns exist.

Within the second part of this work the splicing microarray was transferred to *CYP2B6.* Variant splicing patterns around exon 4 could be observed, which lead to the mechanistic investigation of the observed lower expression phenotype of allele *CYP2B6*6.* The patterns of alternative splicing were studied in detail and the molecular mechanism of the functional impairment allele**6* was clarified, solving a long-existing and important pharmacogenetic problem (Hofmann et al., 2008; Zanger et al., 2007; Zanger and Hofmann, 2008).

4.1 Implementation of the microarray

To produce the microarrays, epoxy-coated slides, 3'-or 5'-aminolink modified oligonucleotides and a spotting buffer, which included DMSO were used. This buffer was used as it produced uniform spot shapes, allowed long spotting rounds with low evaporation rates at 54% humidity (Grimm et al., 2004). As it is known that there might exist variations resulting from the spotting process, 4 different pins were used and always three replicate spots were spotted with the same pin as suggested in a previous publication (Grimm et al., 2004). Running duplicate or triplicate experiments controlled reproducibility between hybridizations.

Overall reduced signals from exon 9 to 1 was seen, but could not be entirely explained, as no difference in GC content, Tm or oligonucleotide length was present. This trend was observed after hybridization of target DNA or cRNA generated by PCR amplification or linear reverse transcription with a gene specific *CYP2D6*-T7 primer placed at the 3'end of exon 9. In the latter case the difference might be explained by the preferential reverse transcription of the 3'end of the gene. But that can't be the only explanation as the bias was also observed with target DNA generated by conventional PCR. More likely, this bias within the signals may be due to the incorporation of Cy3 or aminoallyl modified nucleotides during amplification or transcription. Modified nucleotides are bigger compared to unmodified nucleotides, which will be incorporated preferentially by the polymerase. This may lead to a bias to more Cy3 molecules in the 5'end of the transcribed/ amplified gene and less towards the 3'end.

So far unreported improvement of the protocol was the discovery that addition of SDS in the hybridization buffer results in up to 4 times higher signal intensities and no visible increase of background. It can be speculated that SDS enhances the immobilization efficiency by reducing secondary structures, this way resulting in better accessibility of the target to the probe. However it remains unclear how that might occur mechanistically as both SDS and the oligonucleotides are negatively charged.

The reproducibility is partial critical in the development of microarrays. Numerous factors can affect the hybridization results, starting from manufacturing of the array, the amplification of the labeled target DNA or RNA and different hybridization efficiency depending on the used probes (Bodrossy, 2003). The current state of the

art provides 5%-10% variations in signal among replicate array elements and 10-30% variation among corresponding array elements on different self manufactured cDNA microarrays for expression analysis (Stears et al., 2003; Yue et al., 2001). The standard error of the developed splicing microarray between double or triplicate experiments was between 0 and ±16.8%. So the standard error was in the range as reported for this technology. Reproducible and comparable results were achieved with the designed splicing microarray. The whole spotting and hybridization process was optimized with the result that the standard errors rarely exceeded 5% and never exceeded 17%. High standard errors could be explained in most cases by dust particles at the spot, or problems during the drying process after hybridization, which led to misidentification of some spots. When this happened the spots were excluded.
The sensitivity limit of the developed microarray was determined at 5 ng target DNA (corresponds to ~ 0.01 pmol) that is much better in comparison to published splicing arrays (detection limit of 3 pmol) (Fehlbaum et al., 2005). The 30 times higher sensitivity might be explained by the chosen narrow criteria of probe design, the individual non automated design and validation of each probe. Furthermore it was found that the usage of hybridization buffer including SDS showed up to 10 times higher intensity values compared to standard hybridization buffers. The detection limit was measured with Cy3-labelled target DNA. Usage of cRNA as target might reduce the sensitivity slightly but the array can be considered as very sensitive.

4.2 Platform for analysis of alternative splicing

Alternative splicing arrays are so far commercially based on two technology platforms – the Affymetrix gene technology array based on 25 nt long exon probes and intron control probes and the Agilent/ ExonHit platform with usage of exon and junction probes based on 24 nt probes (Fehlbaum et al., 2005). Both are impressive platforms including probes for all genes of the human genome or in the case of Agilent/ ExonHit for a huge number of genes.

Both systems are using a fixed probe length – what makes it impossible to design probes with identical melting temperature, GC-content or which show no distracting secondary structure (Cuperlovic-Culf et al., 2006). The low oligo length (24 and 25 nt) compared to optimal length of exon probes (50-60 nt) and junction probes validated by Castle et al. (35-40 nt) is possibly leading to low sensitivity (Castle et al., 2003; Srinivasan et al., 2005). Affymetrix arrays try to compensate this by using multiple probes and furthermore adding one mismatch probe for each probe. Developing an alternative splicing array for *CYP2D6*, *CYP2D7* and *CYP2B6* optimal probe length was used in this study between 29-51 nt closely to the guideline rules above.

Most publications focus on the exon boundaries only exon and/ or junction probes. With this probe design mainly known splice forms or alternative cassette exons can be detected (Johnson et al., 2003; Nagao et al., 2005; Fehlbaum et al., 2005; Kapranov et al., 2002; Kampa et al., 2004). The selected probe design based on 5 different probes in this study enables the detection of theoretically over 70% of alternative splicing events including cassette exons, intron retention and alternative 3' or 5'prime sites (if more than 3 base pairs are exchanged) (Watahiki et al., 2004). Recently it was described that junction probes are expected to bring specificity to discriminate a mismatch of 3 base pairs like in NAGNAG insertion-deletions (Fehlbaum et al., 2005).

It was previously reported that some junction probes are not informative due to constitutive high or non-hybridization signal (Nagao et al., 2005). Within this study this was observed with some junction probes. Changing the oligo-dT aminolink from the 3' to the 5' end solved the problem in most cases. Probably because sterical hindrance is minimized and specific hybridization is enhanced. It was mentioned before that the design of junction probes should be improved, especially where splice junctions are difficult to discriminate due to similarity between alternative exons, or

other structures near the splice junctions (Srinivasan et al., 2005). In this study the same problem of cross hybridization was observed for some probes and solved by redesign of some junction, jump-junction and intron-exon probes by extending the probe on one side, what increased the specificity (figure 7). This demonstrated the importance of a non-automated probe design with subsequent evaluation.

To analyze the chosen probe set a new analysis method was developed and evaluated. The data quantification is based on probe sets in which the intensity of individual probes are compared by calculating the relative intensity I_{rel} between different isoforms. This method allows comparing semiquantitatively the expression levels of splice variants within one sample and in comparison of two samples.

Within the evaluation process known DNA fragments (either lacking or including intron regions) were hybridized. After optimization and evaluation of the probes the expected signals for the known fragments were achieved with high specificity and sensitivity for both *CYP2D6* and *CYP2B6.* During evaluation it demonstrated that the developed splicing microarrays can discriminate if intronic regions are included or sliced.

4.3 Alternative splicing within *CYP2D6*

As the evaluation process with known DNA fragments showed high sensitivity and specificity of the expected signals and low background, the next step was to test the developed splicing microarrays with reverse transcribed and labelled RNA of *CYP2D6*1/*1* and *CYP2D6*41/*41* samples. These two genotypes were chosen as it had been previously shown in a recombinant approach as well as in genotyped liver samples that the 2988G>A SNP, the key mutation of allele **41,* leads to decreased protein expression and activity due to enhanced alternative splicing (Raimundo et al., 2004; Toscano et al., 2006). Samples genotyped positively for 2988G>A exhibit decreased amounts of the normal transcript and substantially increased amounts of a splice variant lacking the entire exon 6. This alternative transcript harbours a stop codon at position 291 of the mature mRNA and can thus not contribute to functional CYP2D6 protein. Thereby, the 2988G>A shifts the balance between the functional and a non-functional splice product towards the non-functional variant (Toscano et al., 2006). In contrast samples with **1* or **2* alleles expressed almost exclusively the normal transcript.

The probe signals within *CYP2D6*1/*1* wildtype samples showed the full functional mRNA transcript to be the most common one. Within samples **1/*1* all other signals were below 15% compared to the highest signals, which were seen for the junction probes. It is not known if the small signals in **1/*1* for exon-intron, intron-exon and jump-junction probes, which were below 15% (compared to the signal intensity of the junction probes) were achieved due to crossybridizations or might be explained as high specific signals showing the percentage of biological noise, which occurs during the highly complex splicing process. Therefore only signals higher than 20% were regarded as indicators for potential aberrant splicing events.

Hybridizing samples genotyped positively for 2988G>A (*CYP2D6*41*) on the *CYP2D6* splicing microarray showed equal amounts of an alternative transcript with skipping of exon 6 compared to the normal transcript. In contrast hybridized samples carrying genotype *CYP2D6*1* or **2* showed 5 times higher amounts for the functional transcript compared to the splice variant lacking exon 6. This result reflects the finding of Toscano et al. mentioned above and demonstrates the capacity of the developed splicing microarray for studying exon skipping events (Toscano et al., 2006).

The transcript diversity monitored with the splicing platform in samples with genotype *CYP2D6*1/*1* or *CYP2D6*2/*2* showed similar patterns and no increased alternative splicing. Transcript patterns within the different alleles **1*, **2* and **41* were reproducible in different samples, thus demonstrating allele specific splicing patterns.
Studying the splicing patterns in the whole gene from exon 1-9 in detail it was observed that additional transcript variants occurred at enhanced levels in RNA from *CYP2D6*41* carriers especially total or partial retention of intron 5 and 6. Intronic retention seems to be doubled compared to **1* and was verified by sequencing. The signal for intron 1 was increased but sequence verification failed in this case. Interestingly all intronic splice variants were described before but were not noted to be allele specific (table 1) (Gaedigk et al., 2005). As all observed splicing variants were described before and were specially seen to be enhanced only in allele **41* samples, these results suggest indeed that allele **41* is associated with additional splice variants.
RT-PCR splicing analysis is dependent on amplifiable regions and is limited by the choice of the primers – the whole transcript diversity can only be studied with splicing microarrays. Putting the result together the explanation of the **41* IM phenotype is possible not only by the strongly enhanced transcript lacking exon 6 but also as a sum of the whole transcript diversity within this allele. This might be either explained by additional unknown intronic mutations linked to 2988G>A or by a multiple effect of this mutation on alternative splicing.
The *CYP2D6*4* allele is known to inactivate the splicing acceptor at the exon 4 and to activate a new cryptic site one nucleotide downstream of the natural acceptor. This leads to a premature termination codon (PTC) during transcription, which is known to stop the translation process and to trigger nonsense mediated decay (NMD). NMD is an RNA surveillance mechanism, which leads to degradation of the wrongly spliced mRNA (Cartegni et al., 2002). After linear transcription and hybridization of *CYP2D6* containing haplotype **4/*4* signals were too low for analysis in all but one sample. It is possible that most of the *CYP2D6* mRNA is degraded at the 3' side due to NMD and will be not linearly transcribed. For the sample showing analyzable hybridization signals it was visible that within exon 1-4 normal splicing (highest signal for junction probes) is most common and from there on introns 5-8 are fully or partially retained in up to 50% of the transcripts. This cannot be explained by the **4* key mutation

1846G>A in intron 3 but possibly by additional mutations within the various **4* alleles (**4A-*4N*), or so far unknown intronic mutations.

With the developed splicing microarray it was possible to verify the known splicing events within *CYP2D6.* This confirmed the validity of the used probe design and analysis method. Additionally it was seen that the transcript patterns are not only defined by one or two transcripts per gene. Splicing is typically not an all or none phenomenon (Gardina *et al.*, 2006). Each well defined allele can have its own splicing patterns. The sum of transcript isoforms and the amount of mRNA expression defines the phenotype. For splicing analysis and correct interpretation it is important to genotype the gene and to assign the allele. It is possible that some of the annotated transcripts belong to the biological noise, which is part of the complex expression programs and as such may be functional unimportant.

4.4 Gene and pseudogene specific splice detection of *CYP2D6* and *CYP2D7*

A major problem for the interpretation of data from splicing microarrays is the existence of pseudogenes. By looking at the selective array-probes in detail it is apparent that in so far published splicing microarrays no attention was paid on pseudogenes, especially for pseudogenes corresponding to drug metabolizing enzymes and transporters. If no attention is being paid to pseudogenes (no gene specific transcription or probes are used) the transcripts of the pseudogene would hybridize to the gene probes (due to high homology). If an alternative splicing variation from pseudogenes is annotated to its functional gene this can lead to misinterpretation of the amount of mRNA transcripts and amounts of translated protein. Therefore, one aim of this work was to develop gene/ pseudogene differentiating microarray detection methods.

The highly polymorphic *CYP2D* gene locus contains the active *CYP2D6* and two pseudogenes *CYP2D7* and *CYP2D8*. *CYP2D8* has accumulated numerous deleterious mutations and is non functional. *CYP2D7* is expressed in liver at comparable levels as *CYP2D6* (Endrizzi et al., 2002). In comparison to *CYP2D6*, *CYP2D7* carries one single insertion at position 137 in exon 1 causing a PTC. *CYP2D7* shares 97% sequence homology to *CYP2D6* (Heim and Meyer, 1992).

Both the *CYP2D6* gene as well as the pseudogene *CYP2D7* is alternatively spliced, what complicates the study of *CYP2D6*. 13 different splicing transcripts were described within *CYP2D6* and *CYP2D7* (table 1).

The here described splicing platform is the first where the transcripts of the gene and the pseudogene can be monitored individually. This was achieved by using specially designed gene specific primers including a 3'T7 promoter for linear reverse transcription of cRNA. In addition the gene and pseudogene specific results were achieved by applying a probe design specially adapted to the gene/ pseudogene structures and special pseudogene probes were designed and added if possible.

Analyzing specific *CYP2D7* cRNA it was visible that higher amounts of full or partial intron 5 and 6 retention occurred within *CYP2D7* compared to *CYP2D6*. This fits with the findings from Gaedigk et al. (2005) who also found a high amount of intron retention within *CYP2D7* using a RT-PCR strategy. Transcripts with lack of exon 6 (found to be enhanced in IM allele **41*) were never detected in *CYP2D7*. This supports the finding that the exon 6 splicing event within *CYP2D6*41* can be

explained by the intron mutation 2988G>A. This mutation is not existent within the *CYP2D7* reference sequence. This observation further validated the specific functionality of the developed microarray.

4.5 Alternative splicing within *CYP2B6*

CYP2B6 is a polymorphic gene showing high variability in protein and activity (Lang et al., 2001; Desta et al., 2007; Hofmann et al., 2008). Part of the variation is associated to genotype. Following the discovery that CYP2B6 is the major enzyme for oxidative metabolism of efavirenz (Ward et al., 2003), a number of pharmacogenetics studies with HIV-infected individuals have convincingly shown that homozygosity for *CYP2B6*6* predicts elevated plasma levels of the drug and increased risk for neurotoxicity (Tsuchiya et al., 2004; Haas, 2006; Owen et al., 2006; Rotger et al., 2007; Telenti and Zanger, 2007). Homozygotes for the **6* allele express only about 50% or less CYP2B6 protein in human liver microsomes, compared to homozygotes of the reference **1* allele (Lang et al., 2001; Desta et al., 2007). Microsomal activities towards suitable probe drugs were also found to be decreased to similar extent as shown for S-mephenytoin N-demethylation (Lang et al., 2001), as well as bupropion and efavirenz hydroxylation (Desta et al., 2007). Functional analysis of recombinantly expressed variants of the involved amino acid variants (Q172H, K262R, or both) revealed some controversial results. When expressed as N-terminally modified enzyme in *E. coli*, 7-ethoxycoumarin O-deethylase activity of the Q172H mutant (allele **9*) did not follow Michaelis-Menten type kinetics but displayed sigmoidal kinetics with increased turnover (Ariyoshi et al., 2001). In COS-1 cells the Q172H + K262R double variant expressed at normal levels compared to the wild type but bupropione hydroxylation was catalyzed less efficiently (Rotger et al., 2007; Jinno et al., 2003). So far the molecular mechanism behind *6 was not clarified.

Eight splice variants are known in the literature but allele specific transcript patterns have not been studied yet. With the *CYP2B6* splicing microarray it should be investigated if the *CYP2B6*6* phenotype might be connected to a different allele specific transcript patterns compared to *CYP2B6*1.* The splicing platform, which showed successful detection of aberrant splicing and detection of transcript patterns within *CYP2D6* and *CYP2D7*, was therefore transferred to *CYP2B6*. *CYP2B6* was long time neglected because of its low abundance of the total P450 liver enzymes (about 4%). This was indeed problematic for the splicing platform as the signals of hybridized samples of *CYP2B6* cRNA were much lower and only in some samples high enough to achieve useful results.

Samples from patients with *CYP2B6*6/*6* and **1/*1* genotype studied on the *CYP2B6* splicing array showed indeed different hybridization pattern with enhanced lack of exon 4 (SV8) and partial intron 3 retention (SV2-SV6). Aberrant splicing of partial intron 3 (SV2-SV5) was the first alternative splicing observation made in *CYP2B6* by Miles et al. (1989). However the lack of exon 4 was never described before. Due to these results achieved with the splicing array, RT-PCR analysis was performed and confirms the new variant lacking exon 4 (SV7, table 4).

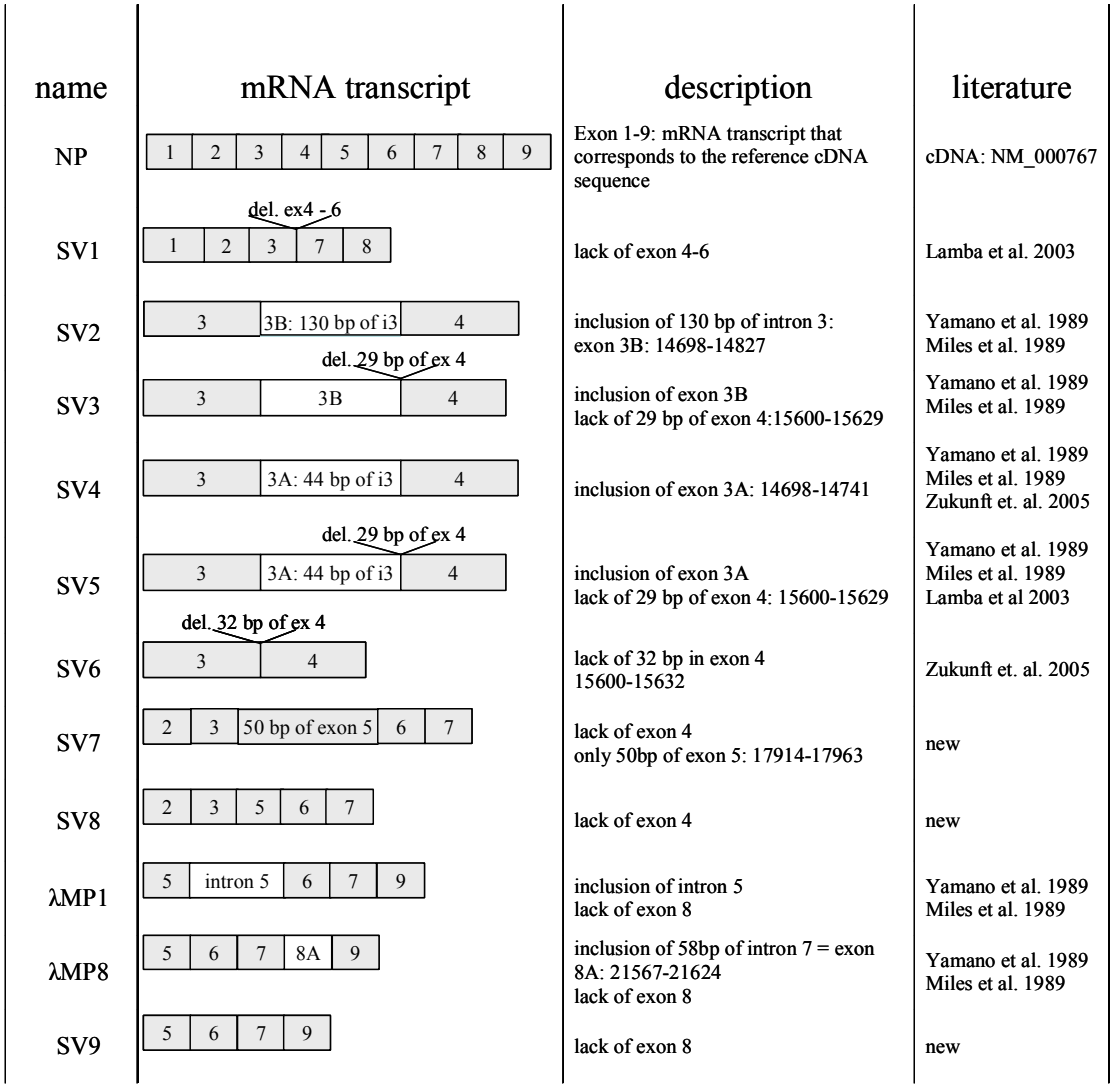


Table 4: Compilation of transcript variants of *CYP2B6*

Shaded boxes represent exonic and white boxes intronic sequences. If possible, nomenclature of spliced transcripts were used as described before (Lamba et al., 2003; Miles et al., 1989). SV6 was newly designated and SV7 as well as SV8 are novel splice variants not described before.

Furthermore using RT-PCR it showed that the region between exon 3-7 produces a hudge number of different alternative splice variants (table 4). A new splice variant SV9 lacking exon 8 was observed by RT-PCR. This variant was not observed with the junction probe from exon 7 to 9. A possible explanation for the unspecificity of this junction probe might be unknown mutations within this region or a processing error in setting up the spotting plate for the *CYP2B6* microarray.
Additionally by RT-PCR the new splice variant SV8 and the previously described splice variant SV1 were observed but both were missed by the *CYP2B6* splicing array. The splicing array observed lacking of exon 4 but not the SV1 lacking exon 4-6 or SV8 lacking exon 4 and partial exon 5 as with the used probe structure only lacking of one but not 2 or 3 exons in a row could be detected. These results point out that for further array development the probe selection should be adjusted by adding additional jump-junction probes stretching from one exon to the third, fourth, fifth and last exon.

Subsequent microarray experiments, confirmatory RT-PCR analysis and quantification of products using the Agilent Bioanalyzer lead to an highly interesting observation: the existence of a so far undetected allele specific bias was observed (Hofmann et al., 2008). Most interesting it was found that the previously described SV1 lacking exon 4-6 (Lamba et al., 2003) was strongly enhanced in samples with genotype **6/*6* but hardly detectable in **1/*1* samples. The major transcript detected in **1/*1* samples was the normal splice product whereas in **6/*6* samples it was the SV1 variant (table 3 and figure 20). The oberserved splicing pattern of *CYP2B6*1* compared to *CYP2B6*6* lead to the idea that the reduced protein and activity observed in previous studies might be explained due to alternative splicing.

4.6 Elucidation of the mechanism behind *CYP2B6*6*

Analysis of existing protein expression data from a collection of 235 human liver samples confirmed significantly reduced protein and activity in individuals with **6/*6* compared to **1/*1* genotype, as previously described for a smaller sample set (Desta et al., 2007). Interestingly, the specific activity for bupropion hydroxylation was not reduced in liver samples for **1/*6* and **6/*6* indicating that the amino acid changes do not have major influence on enzyme activity as shown in the PhD thesis of Blievernicht 2008 (Hofmann et al., 2008) (figure 34).

Based from the microarray analysis and subsequent RT-PCR analysis it was speculated that the effect of enhanced alternative splicing should be visible by reduced amount of full functional mRNA (NP), which would result in reduced protein expression thus explaining the **6* phenotype. The *CYP2B6* mRNA levels in the liver sample collection were quantified with a previously developed TaqMan assay (Zukunft et al., 2005). The amount of mRNA levels in relation to genotype, revealed significantly reduced levels for **6/*6* and for **1/*6.* This result is in concordance with the microarray and RT-PCR analysis (figure 22 and 34).

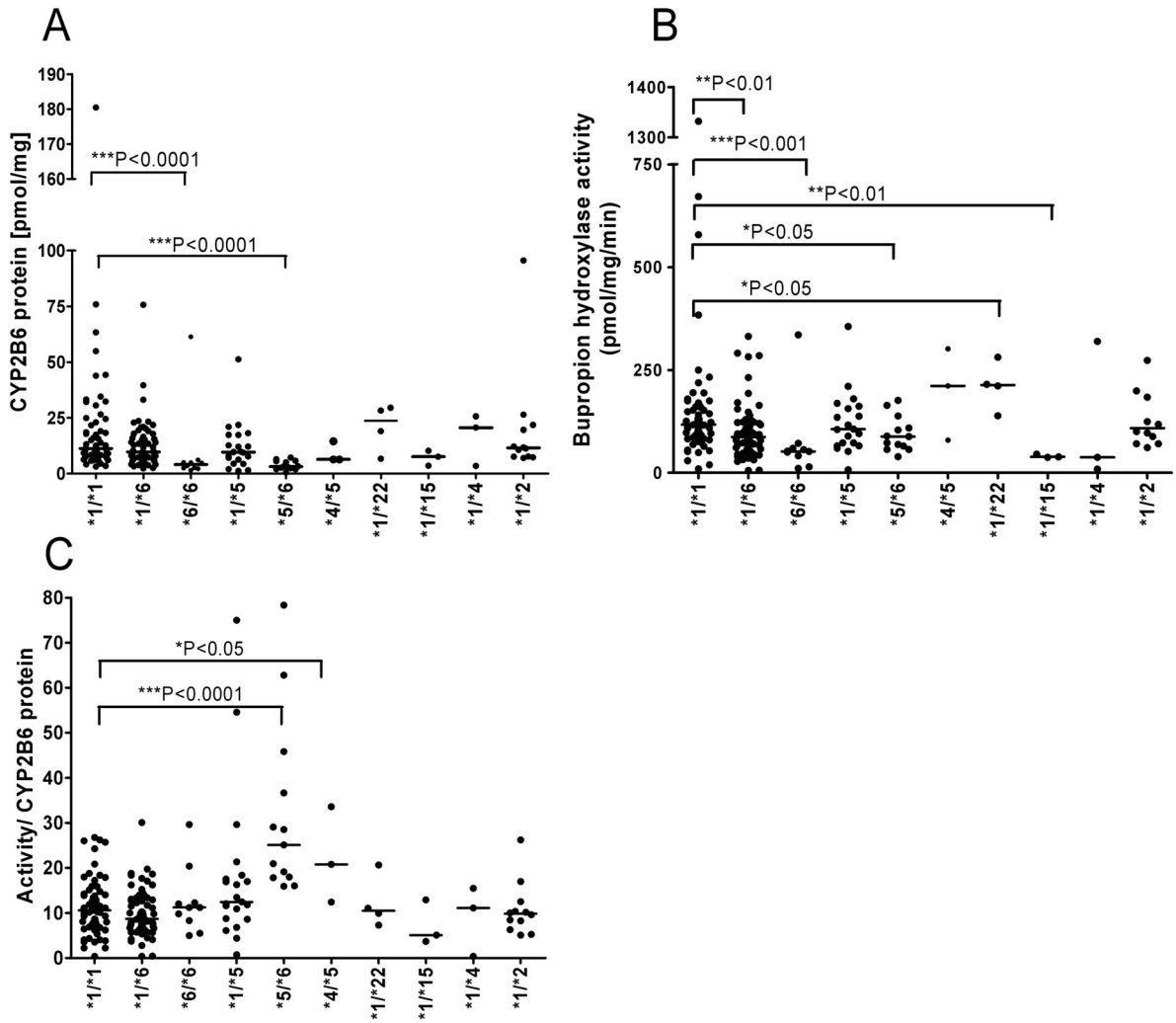


Figure 34: Relationship between genotype and CYP2B6 protein expression, microsomal bupropion (50 µM) hydroxylase activity or specific activity in human liver samples. Blievernicht J.K. studied these results in her PhD thesis (2008).

A: Scatterplots depicting CYP2B6 protein expression derived from western blot analysis shown in pmol/mg microsomal protein for various genotypes.

B: Microsomal bupropion (50 µM) hydroxylase activity as a marker for CYP2B6 activity in relation to various genotypes.

C: Specific activity calculated as microsomal activity divided by specific CYP2B6 protein content within each liver sample in relation to various genotypes. Some genotype groups are not shown due to insufficient number of individuals. Statistical significance is indicated by brackets pointing at the two groups compared by Mann-Whitney test.

Surprisingly mRNA reduction was not described within *CYP2B6*6/*6* in a previous study (Hesse et al., 2004). Interestingly, comparing the TaqMan assay used within this study with the assay used by Hesse et al., it can be expected that the latter also detects splice variant SV1, because they placed 7 hybridisation probes between exons 1 to 4 (Hesse et al., 2004). This explains why no correlation was found between genotype and transcript level in the former study.

In contrast to that the assay used within this study definitely does not detect SV1, SV7 and SV8 as primer and probes are all placed within exon 4 and 5. Therefore the amount of the mRNA corresponds to the normal product and thus correlates to genotype (figure 35) (Hofmann et al., 2008).

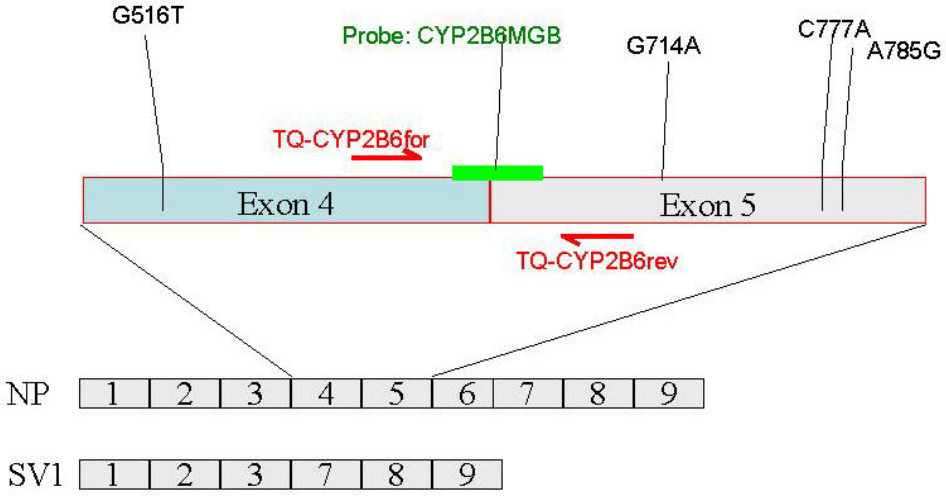


Figure 35: *CYP2B6* mRNA quantification by specific TaqMan real-time RT-PCR assays

The assay showed specificity for all mRNA transcripts which include exons 4 and 5: for example NP; and no signal for transcripts lacking exon 4 or exon 4-6: SV1, SV7 and SV8. Most common mutations within this region are included in the figure.

The *CYP2B6* microarray results, the RT-PCR analysis as well as the mRNA quantifications with the TaqMan assay indicated that the reduced mRNA amount in *CYP2B6*6* might be explained by aberrant splicing. For mechanistic explanation the key mutation inducing alternative splicing needs to be identified.

In a previous study from Lamba et al. several splice variants including one that lacked sequence corresponding to exons 4, 5, and 6 were found in many of their liver samples analyzed (Lamba et al., 2003). The presence of SNP c.516G>T in exon 4 and of g.15582C>T in intron 3 was found to correlate with the appearance of splice

variant SV1. *In silico* analysis of splice sites suggested that mutation g.15582C>T disrupts a cryptic splicing acceptor in intron 3 and c.516G>T in exon 4 may modulate an exonic splicing enhancer. Phenotypic changes were however not unequivocally related to one or the other SNP (Lamba et al., 2003). In this study genotyping of g.15582C>T and correlation analysis with existing mRNA, protein and activity data could not confirm an association between g.15582C>T and hepatic phenotype (figure23). RT-PCR analysis of samples carrying this SNP compared to non-carriers did not reveal any different levels of SV1 (see figure 24).

It seemed as if the splicing acceptors of exon 4 and exon 5 are extremly weak and might be influenced by mutations in the near area. Because the **6* allele includes the two non-synonymous SNPs in exon 4 and exon 5 and possibly additional unknown SNPs in intronic regions, a recombinant minigene approach was applied to unequivocally identify the causal mutation(s) leading to enhanced alternative splicing. The minigenes were constructed to include the entire *CYP2B6* coding region and most of the intronic regions in order to allow analysis not only at the mRNA level but also at the protein and activity level. Indeed, transfection of two different mammalian cell lines resulted in full-length CYP2B6 protein, which was functionally active. These constructs, which differed in only one or two (only for the [**6*] construct) mutations, resulted in strikingly different phenotypes with consistent changes at all levels of gene expression. Whereas the construct representing the **4* allele resulted in essentially unchanged gene expression, the constructs with the c.516G>T mutation, representing alleles **9* and **6*, resulted in high levels of SV1, low levels of normal splice product, low protein levels and low enzyme activity.
These results were only apparent when the plasmids were diluted with control plasmid at least 1000-fold. Transfections of higher concentrations of the plasmids resulted in additional alternative spliced transcripts and reduced amount of the NP. Within a single cell only a limited amount of proteins and snRNPs are available which can be used to build up the multicomponent ribonucleoprotein complex – the splicosome, which is responsible for correct recognizing the splicing acceptor and donor (for example SR-proteins). It can be speculated that transfected cells might overexpress minigene-derived pre-mRNA to such an extent that splicing cannot be properly performed due to limiting amounts of splicing factors. Indeed, at 1000-fold dilution the mRNA patterns of transfected cells corresponded well to those in human

liver samples of the same genotype. Mechanism of the *CYP2D6*41* allele had been investigated with a similar minigene approach, but to this time no dilution of minigene was perfromed (Toscano et al., 2006). Within this study also enhanced levels of aberrant splice products were observed in transfected cells. The additional transcripts were independent of the genotype. Summarizing the above it can be said that, for studying splicing in cell systems the amount of transfected minigene needs to be highly reduced, compared to concentrations, which are commonly used in expression experiments.

It should be added that the effect of the single mutation was even stronger than expected, as it exceeded the changes in expression in liver. A possible explanation of this might be that under *in vivo* conditions in the liver, the penetrance of alternative splicing effects is less than under cell culture conditions.

4.7 In silico analyis and evolutionary considerations of alternative splicing

In silico analysis using the ESE finder (Cartegni et al., 2003; Smith et al., 2006) showed that the presence of c.516G>T in *2B6*6* is predicted to change the scores for binding of certain splice factors to ESE consensus sequences (figure 33) (Zanger and Hofmann, 2008). This probably results in the modulation of splicing events in favour of an alternative process that results in the observed splice variant lacking exons 4, 5 and 6. However involvement of these splicing factors in determining alternative splicing of *CYP2B6* remains speculative, as their precise functions can not yet be predicted and additionally from this model it can not be read out why three and not only one exon is skipped. It can be speculated that in addition so far unknown intronic mutations might exist within intron 6. All approaches used in this study sum up to the conclusion that the exon 4 SNP c.516G>T leads to lower *CYP2B6* expression via quantitative modulation of splicing through skipping of exon 4-6.

Finally, evolutionary considerations revealed some interesting aspects. There is evidence that *CYP2B6* and the *CYP2B7* pseudogene are derived from a gene duplication event (Hoffman et al., 2001; Wang et al., 2003a) interestingly the key mutations defining *CYP2B6*6* (i.e. c.516G>T and c.785A>G) are both present on the reference sequence of *CYP2B7* (haplotype TG) whereas the orthologous rodent genes *CYP2B1/2* and *cyp2b9/10* correspond to the GG haplotype. Although no polymorphism data are available for *CYP2B7* it can be supposed that *CYP2B7* is invariable at these two positions, i.e. of "**6*" genotype. This would mean that the original *CYP2B6* gene derived from *2B7* by gene duplication was the *CYP2B6*6* allele and that today's *CYP2B6 *1* was formed by sequential point mutations. It would be interesting for future investigations to study splicing of the pseudogene *CYP2B7*, either *in vitro* using cell culture or using human lung tissue in which expression of *CYP2B7* was reported (Gonzalez et al., 1992) and to analyze if there exist similar splicing patterns as in *CYP2B6*6*. It should be noted that the hypothesis that today's *CYP2B6*1* was formed after gene duplication by sequential point mutations corresponds very well with previous observation from Zukunft et al. (2005) who showed that the key mutation for **22* (-82T>C) in the promoter region can be found in the pseudogene *CYP2B7*, the murine *cyp2b9/10* and chimpanzee orthologous gene CYP2Bch.

The allele **1* codes for an enzyme with higher enzyme activity which might be advantageous it allows more rapid detoxification and clearance of xenobiotics.

The observed complicated splicing patterns certainly increase the variability of *CYP2B6* which is probably an advantage in some circumstances. Dependent on the xenobiotics the human exposes himself; it is either advantage or disadvantage to have a higher or lower CYP2B6 enzyme activity. It remains unknown whether the pronounced aberrant splicing observed in a very common human allele has been - or still is - a target for evolutionary selection.

5 Conclusion

The goal of this project was the development of a microarray platform to detect and further investigate alternative splicing variants in human liver tissue. First a microarray platform was developed for detection of splicing patterns in the drug metabolizing cytochrome P450 enzyme CYP2D6 and in the pseudogene *CYP2D7*. The systematic probe design and optimization of all steps from microarray production to hybridization lead to a sensitivity limit 100 fold better compared to described splicing arrays. The assay enabled the detection of known *CYP2D6* splicing variants and specific splicing patterns were monitored for the most common *CYP2D6* alleles **1*, **2*, **4* and **41*. Semi-quantitative data analysis was possible with a new data quantification method developed for the used alternative probe set. This is the first described array platform which can distinguish between gene and of the highly similar pseudogene. This microarray proved to be a valuable tool for studying alternative splicing and in future it will be interesting to use this tool in different tissue types.

The platform was adaptable to other genes as shown for the second gene *CYP2B6.* Within this gene different allele specific splicing patterns were visible with the splicing platform comparing allele **1/*1* with **6/*6*. This result was verified by comprehensive RT-PCR analysis and this way three new splice variants (SV7, SV8 and SV9) were found in *CYP2B6*. This lead to the assumption that alternative splicing might explain the molecular mechanism of *CYP2B6*6* with decreased expression and function in human liver.

With a molecular approach using recombinant minigenes the mechanism of allele *CYP2B6*6* showing reduced function, was clarified. Data clearly establish that, first, the exon 4 SNP c.516G>T, one of the two linked non-synonymous polymorphisms in the allele**6* is the causal mutation finally leading to reduced function; second, that this mutation is responsible for aberrant splicing resulting in a major splice variant that lacks the region including exons 4 to 6 and the introns in between; and third that this results in severely reduced levels of functional full-length mRNA transcript, protein and activity. This is an important finding because the SNP c.516G>T can now be reliably used in pharmacogenetic studies in various clinical settings including

prediction of drug plasma concentration, toxicity, drug effectiveness and dose adjustment as in the cause of the anti-HIV drug efavirenz.

In conclusion, single nucleotide polymorphisms that lead to aberrant splicing are very common among drug metabolizing cytochromes P450 enzymes. Therefore, aberrant splicing can be regarded as a major mechanism of polymorphic expression of drug metabolizing cytochromes P450. In these gene family alelle specific splicing should be studied in more detail using specific microarry platforms as described in this study and recombinant minigenes to find the causal mutation for alternative splicing variants.

6 Appendix

6.1 Probes used for the splicing microarray

Table 1: Specific probes which were used for fabrication of splicing microarrays.
Column 1: Name of the probes; column 2: their specificity either for *CYP2D6*, *CYP2B6* or its pseudogenes *CYP2D7* or *CYP2B7*; column 3 and 4: position of the probes; column 5: probe sequence with additional 15 thymidine spacer; column 6: modification to bind the oligonucleotide probe on the epoxy surface of the microarray slides.

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-004-E10	*CYP2D6/7*	UTR	Exon 1	ttttttttttttttttggtcacgcgctcggtgtgctgagagtgtcctgcct	C6-Aminolink
H041012-004-F10	*CYP2D6*	Exon 1	Exon 1	ttttttttttttttttaggtgtgtccagaggagcccatttggtagtgaggcaggtatggggctaga	C6-Aminolink
H041012-004-G10	*CYP2D7*	Exon 1	Exon 1	ttttttttttttttttaggtgtgtccagaggagcccagttggtagtgaggcagccatggggctaga	C6-Aminolink
H041012-004-H10	*CYP2D6*	Exon 1	Exon 1	ttttttttttttttttggtagtgaggcaggtatggggctagaagcactggtgcccctgg	C6-Aminolink
H041012-004-A11	*CYP2D7*	Exon 1	Exon 1	ttttttttttttttttggtagtgaggcagccatggggctagaagcactggtgcccctgg	C6-Aminolink
H041012-004-B11	*CYP2D6/7*	Exon 1	Exon 1	ttttttttttttttttgctgcatgtggacttccagaacacaccatactgcttcgaccag	C6-Aminolink
H041012-004-C11	*CYP2D6/7*	Exon 1	Intron 1	ttttttttttttttttccatactgcttcgaccaggtgagggaggaggtcctgga	C6-Aminolink
H041012-004-D11	*CYP2D6/7*	Exon 1	Exon 3	ttttttttttttttttaccatactgcttcgaccaggggtgttcctggcgcgc	C6-Aminolink
H041012-004-E11	*CYP2D6/7*	Exon 1	Exon 2	ttttttttttttttttcatactgcttcgaccagttgcggcgccgcttcgg	C6-Aminolink
H041012-004-F11	*CYP2D6/7*	Intron 1	Exon 2	ttttttttttttttttctgaccctccctctgcasttgcggcgccgcttcgg	C6-Aminolink
H041012-004-G11	*CYP2D6*	Exon 2	Exon 2	ttttttttttttttttctgtgcccatcaccccagatcctgggtttygggcc	C6-Aminolink
H041012-004-H11	*CYP2D7*	Exon 2	Exon 2	ttttttttttttttttctgcgcccatctaccaggtcctgggcttcgggcc	C6-Aminolink
H041012-004-A12	*CYP2D6/7*	Exon 2	Exon 2	ttttttttttttttttctggacgccggtggtcgtgctcaatgggctg	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-004-B12	*CYP2D6*	Exon 2	Exon 2	ttttttttttttttgccgcttcgggg acgtgttcagcctgcagct	C6-Aminolink
H041012-004-C12	*CYP2D6*	Exon 2	Intron 2	ttttttttttttttggccgcgttccc aaggcaagcagcggtgg	C6-Aminolink
H041012-004-D12	*CYP2D7*	Exon 2	Intron 2	ttttttttttttttggccgcgttccc aaggcaagcggcggtgg	C6-Aminolink
Platte 1-F12	*CYP2D6/7*	Exon 2	Exon 3	ttttttttttttttgggccgcgttcc caaggggtgttcctggcg	C6-Aminolink
H041012-004-E12	*CYP2D6/7*	Exon 2	Exon 4	ttttttttttttttygggccgcgttc ccaaggacsscccttygccc y	C6-Aminolink
H041012-004-F12-inver	*CYP2D6*	Intron 2	Exon 3	cgtcccacccccaggggtgtt cctggcgcgctatttttttttttttt	C7-Aminolink
H041012-004-H12	*CYP2D7*	Intron 2	Exon 3	ttttttttttttttcgtcccgccccc aggggtgttcctggcgcgcta	C6-Aminolink
H041012-005-A01	*CYP2D6/7*	Exon 3	Exon 3	ttttttttttttttcgctatgggccc gcgtggcgcgagcagagg	C6-Aminolink
H041012-005-B01	*CYP2D6/7*	Exon 3	Exon 3	ttttttttttttttrtstccaccttgcg caacttgggcctgggcaaga agtcg	C6-Aminolink
H041012-005-C01	*CYP2D6/7*	Exon 3	Exon 3	ttttttttttttttaggaggcygcct gcctttgtgccgccttcgcc	C6-Aminolink
H041012-005-D01	*CYP2D6/7*	Exon 3	Intron 3	ttttttttttttttccttcgccracca mkcyrgtgggtgatgggcag aa	C6-Aminolink
H041012-005-E01	*CYP2D6/7*	Exon 3	Exon 5	ttttttttttttttgccttcgccracc amkcyrgtgctgaatgctgtc cc	C6-Aminolink
H041012-005-F01	*CYP2D6/7*	Exon 3	Exon 4	ttttttttttttttccttcgccracca ctcyrgacgccccttygcc	C6-Aminolink
H041012-005-H01	*CYP2D6/7*	Intron 3	Exon 4	ttttttttttttttgcatctcccacc cccargacsscccttygccc	C6-Aminolink
H041012-005-A02	*CYP2D6/7*	Exon 4	Exon 4	ttttttttttttttttttygccccaacg gyctcttggacaaagccgtga gcaacgt	C6-Aminolink
H041012-005-C02	*CYP2D6/7*	Exon 4	Exon 4	ttttttttttttttgcctccctcacct gcgggcgccgcttcgagtac gacgaccct	C6-Aminolink
H041012-005-D02	*CYP2D6/7*	Exon 4	Exon 4	ttttttttttttttcrcttcctcaggct gctggacctagctcaggaggr ayygaaggaggagt	C6-Aminolink
Platte2-E02	*CYP2D6*	Exon 4	Intron 4	ttttttttttttttctttctgcgcgag gtgcggagcgagagac	C6-Aminolink
H041012-005-F02	*CYP2D7*	Exon 4	Intron 4	ttttttttttttttagtcgggcttcct gcgcgaggtgcggagcaag ggtc	C6-Aminolink
H041012-005-G02	*CYP2D6*	Exon 4	Exon 6	ttttttttttttttgtcgggctttctgc gcgaggccaaggggaaccc	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
				tgaga	
H041012-005-H02	*CYP2D7*	Exon 4	Exon 6	ttttttttttttttgtcgggcttcctgcgcgaggccaaggggagccctgaga	C6-Aminolink
Platte2-A03	*CYP2D6/7*	Exon 4	Exon 5	ttttttttttttttcttyctgcgcgaggtgctgaatgctgtccccg	C6-Aminolink
H041012-005-A03	*CYP2D6/7*	Exon 4	Exon 5	ttttttttttttttgtcgggcttyctgcgcgaggtgctgaatgctgtcccc	C6-Aminolink
H041012-005-B03	*CYP2D6/7*	Intron 4	Exon 5	ttttttttttttttccccgttctgtctggtgtaggtgctgaatgctgtcccg	C6-Aminolink
H041012-005-C03	*CYP2D6/7*	Exon 5	Exon 5	ttttttttttttttagcgytgkctggcaaggtcctacgcttccaaaaggctttcctgac	C6-Aminolink
H041012-005-D03	*CYP2D6/7*	Exon 5	Exon 5	ttttttttttttttgagcacaggatgacctgggacccagcccagccm	C6-Aminolink
H041012-005-G03	*CYP2D6/7*	Exon 5	Exon 5	ttttttttttttttmccccgagacctgactgaggccttcctggcagagatggagaag	C6-Aminolink
H041012-005-B04	*CYP2D6/7*	Exon 5	Intron 5	ttttttttttttttcctggcagagatggagaaggtgagagtggctgccacggtg	C6-Aminolink
Platte2-E03	*CYP2D6/7*	Exon 5	Exon 7	ttttttttttttttcctggcagagatggagaaggccgtgtccaacaggagatc	C6-Aminolink
H041012-005-F04	*CYP2D6/7*	Exon 5	Exon 6	ttttttttttttttcctggcagagatggagaaggccaaggggarccctgag	C6-Aminolink
H041012-005-H04	*CYP2D6*	Intron 5	Exon 6	ttttttttttttttgctctcggccctgctcaggccaaggggaaccctgag	C6-Aminolink
H041012-005-A05	*CYP2D7*	Intron 5	Exon 6	ttttttttttttttcctctcggccctgctcaggccaaggggagccctgag	C6-Aminolink
H041012-005-B05	*CYP2D6/7*	Exon 6	Exon 6	ttttttttttttttaggggarccctgagagcagcttcaatgatgagaacctgygcmtagtgg	C6-Aminolink
H041012-005-C05	*CYP2D6*	Exon 6	Exon 6	ttttttttttttttttgacctgttctctgccgggatggtgaccacctcgacca	C6-Aminolink
H041012-005-D05	*CYP2D7*	Exon 6	Exon 6	ttttttttttttttttaacctgttccttgccgggatggtgaccaccttgacca	C6-Aminolink
H041012-005-E05	*CYP2D6/7*	Exon 6	Exon 6	ttttttttttttttcctggggcctcctgctcatgatcctacmycyrgatgtgcagc	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-005-F05	*CYP2D6*	Exon 6	Intron 6	ttttttttttttttctacmtcyrgatgtgcagcstgagcccatctggaaacagt	C6-Aminolink
H041012-005-G05	*CYP2D7*	Exon 6	Intron 6	ttttttttttttttctacacctggatgtgcagcgtgagcccagctgggcccaag	C6-Aminolink
H041012-005-H05	*CYP2D6/7*	Exon 6	Exon 7	ttttttttttttttctacmtcyrgatgtgcagcggaacgacactcatcacc	C6-Aminolink
H041012-005-A06	*CYP2D6*	Exon 6	Exon 7	ttttttttttttttttacmtcyrgatgtgcagcgccgtgtccaacaggaga	C6-Aminolink
H041012-005-B06	*CYP2D7*	Exon 6	Exon 7	ttttttttttttttttacacctggatgtgcagctccgtgtccaacaggaga	C6-Aminolink
H041012-005-D06	*CYP2D6/7*	Exon 7	Exon 7	ttttttttttttttcaacaggagatcgacgacrtgatagggcaggtgsggcga	C6-Aminolink
H041012-005-E06	*CYP2D6/7*	Exon 7	Exon 7	ttttttttttttttgatgggtgaccaggctcacatgccctacaccactgccgtgatt	C6-Aminolink
H041012-005-F06	*CYP2D6/7*	Exon 7	Exon 7	ttttttttttttttgacccatatgacatcccrtgacatcgaagtacagggcttccgcatcccta	C6-Aminolink
H041012-005-G06	*CYP2D6/7*	Exon 7	Intron 7	ttttttttttttttgcttccgcatcccctaaggtaggcctggcgccctcc	C6-Aminolink
H041012-005-H06-inver	*CYP2D6/7*	Intron 6	Exon 7	ggggacgcatgtctgtccaggccgtgtccaacaggagatctttttttttttttt	C7-Aminolink
Platte2-H06	*CYP2D6/7*	Exon 7	Exon 9	ttttttttttttttttccgcatcccctaaggccgccgtgcatgcct	C6-Aminolink
H041012-005-A07	*CYP2D6/7*	Exon 7	Exon 8	ttttttttttttttcagggcttccgcatccctaagggaacgacactcatcaccaa	C6-Aminolink
H041012-005-B07	*CYP2D6/7*	Intron 7	Exon 8	ttttttttttttttaccctgcatctyctgcccagggaacgacactcatcacc	C6-Aminolink
H041012-005-C07	*CYP2D6/7*	Exon 8	Exon 8	ttttttttttttttttcctgcctttctcagcaggtgcctgtggggagccccg	C6-Aminolink
H041012-005-D07	*CYP2D6/7*	Exon 8	Exon 8	ttttttttttttttccgtctggragaagcccttccgcttccaccccsaa	C6-Aminolink
H041012-005-E07	*CYP2D6/7*	Exon 8	Exon 8	ttttttttttttttcyggatgcccaggccactttgtgaagccggaggc	C6-Aminolink
H041012-005-F07	*CYP2D6/7*	Exon 8	Intron 8	ttttttttttttttttcctgcctttctcagcaggtgcctgtggggagccccg	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-005-G07	*CYP2D6/7*	Exon 8	Exon 9	ttttttttttttttttcctgcctttctcagcaggccrccgtgcatgcctc	C6-Aminolink
H041012-005-H07	*CYP2D6/7*	Intron 8	Exon 9	ttttttttttttttttctcccctccccacaggccrccgtgcatgcct	C6-Aminolink
H041012-005-A08	*CYP2D6/7*	Exon 9	Exon 9	ttttttttttttttttcggggagccccctggcccgcatggagctcttcctctt	C6-Aminolink
H041012-005-B08	*CYP2D6*	exon9	Exon 9	ttttttttttttttttcagccaccatggtgtctttgctttcctggtgasccccatccccctat	C6-Aminolink
H041012-005-C08	*CYP2D7*	Exon 9	Exon 9	ttttttttttttttttcagccactctcgtgtcgtcagctttctggtgacccccatccccctac	C6-Aminolink
H041012-005-D08	*CYP2D6/7*	Exon 9	Exon 9	ttttttttttttttttcccgctagaatgggtacctagtccccagcctgctccctagccagaggct	C6-Aminolink
H041012-005-E08	*CYP2D6/7*	Exon 9	UTR	ttttttttttttttttaaagcaatgtggtagttccaactcgggtcccctgctcacg	C6-Aminolink
Platte 1-A01	*CYP2B6*	UTR	Exon 1	agggcagtcagaccaggaccatggaactcagcgtcctcctcttttttttttttttt	C7-Aminolink
H041012-004-C01	*CYP2B6/7*	Exon 1	Exon 1	ttttttttttttttttcctcctcttccttgcactcctcacaggmctcttgctactcctggttcwgy	C6-Aminolink
H041012-004-D01	*CYP2B6*	Exon 1	Exon 1	ttttttttttttttttgactcttgctactcctggttcwgygccaccctaacacccatgaccgc	C6-Aminolink
H041012-004-F01	*CYP2B6*	Exon 1	Exon 1	ttttttttttttttttggaaaccttctgcagrtggatagaagaggcctactcaaatcctttctgagg	C6-Aminolink
H041012-004-H01	*CYP2B6/7*	Intron 1	Exon 1	ttttttttttttttttggcctactcaaatcctttctgagggtaagacacagacgaatggggtctg	C6-Aminolink
Platte 1-H01	*CYP2B6*	Exon 1	Intron 1	ttttttttttttttttctcaaatcctttctgagggtaagacacagacgaatggggtct	C6-Aminolink
Platte 1-A02	*CYP2B6*	Exon 1	Exon 3	cctactcaaatcctttctgaggcatgctctttgccaatggaaacttttttttttttttt	C7-Aminolink
H041012-004-C02	*CYP2B6*	Exon 1	Exon 2	ttttttttttttttttaggcctactcaaatcctttctgaggttccgagagaaatatggggacgt	C6-Aminolink
H041012-004-D02	CYP2B6/7	Intron 1	Exon 2	ttttttttttttttttccctggtgtggatgtgattggcagttccgagagaaatatggggacgt	C6-Aminolink
H041012-004-E02	CYP2B6/7	Exon 2	Exon 2	ttttttttttttttttttccgagagaaatatggggacgtcttcacggtacacctgggaccsagg	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-004-F02	CYP2B6/7	Exon 2	Exon 2	ttttttttttttttttcccgtggtcatgctgtgtggagtagaggccatacgggagg	C6-Aminolink
H041012-004-G02	CYP2B6	Exon 2	Exon 2	ttttttttttttttttggggaaaaatcgccatggtcgacccattcttccggggatatg	C6-Aminolink
H041012-004-A03	CYP2B6	Exon 2	Intron 2	ttttttttttttttttcccattcttccgggatatggtgagagcctcagaggcactggg	C6-Aminolink
H041012-004-C03	CYP2B6	Exon 2	Exon 4	ttttttttttttttttccattcttccggggatatggggccctcatggacccccac	C6-Aminolink
H041012-004-E03	CYP2B6	Exon 2	Exon 3	ttttttttttttttttcccattcttccgggatatggtgtgatctttgccaatgg	C6-Aminolink
H041012-004-G03	CYP2B6	Intron 2	Exon 3	ttttttttttttttttcacacctccccctgcaccccaggtgtgatctttgccaatg	C6-Aminolink
H041012-004-A04	2B6/7	Exon 3	Exon 3	ttttttttttttttttgccaatggaaaccgctggaaggtgcttcggcgattctctgt	C6-Aminolink
H041012-004-B04	CYP2B6/7	Exon 3	Exon 3	ttttttttttttttttgacttcgggatgggaragcrgagtgtggaggagcggatt	C6-Aminolink
H041012-004-C04	CYP2B6/7	Exon 3	Exon 3	ttttttttttttttttggattcaggaggasgctcagtgtctgatagaggarcttcggaaatccaagg	C6-Aminolink
P1-tube	CYP2B6/7	Exon 3	Intron 4	ttttttttttttttttgarcttcggaaatccaagggtgagtcctgggrgayga	C6-Aminolink
H041012-004-D04	CYP2B6/7	Exon 3	Exon 5	ttttttttttttttttggarcttcggaaatccaaggctgtttgagctcttctctgg	C6-Aminolink
Platte 1-E04	CYP2B6	Exon 3	Exon 4	gcttcggaaatccaagggggccctcatggacccctttttttttttttttt	C7-Aminolink
H041012-004-G04	CYP2B6	Intron 3	Exon 4	ttttttttttttttttgayctgctgcttcttcctaggggccctcatggaccca	C6-Aminolink
H041012-004-A05	CYP2B6/7	Exon 4	Exon 4	ttttttttttttttttrggccctcrtggaccccaccttcctcttccaktccattaccg	C6-Aminolink
H041012-004-B05	CYP2B6/7	Exon 4	Exon 4	ttttttttttttttttcatcatctgctccatcrtctttggaaaacgmttccactaccaagatcaaga	C6-Aminolink
H041012-004-C05	CYP2B6	Exon 4	Exon 4	ttttttttttttttttctgaacttgttctaccagacttttcactcatcagctctgtattcggccag	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-004-E05	CYP2B6	Exon 4	Intron 4	ttttttttttttttagctctgtattcggccaggtcagggagacggagaggga	C6-Aminolink
H041012-004-G05	CYP2B6	Exon 4	Exon 6	ttttttttttttttcagctctgtattcggccaggagaaatccaacgcacaca	C6-Aminolink
Platte 1-A06	CYP2B6	Exon 4	Exon 5	cagctctgtattcggccagctgtttgagctcttctctggttttttttttttttt	C7-Aminolink
H041012-004-C06	CYP2B6/7	Intron 4	Exon 5	ttttttttttttttccacccctttctttcttgcagctgtttgagctcttctctgg	C6-Aminolink
H041012-004-D06	*CYP2B6/7*	Exon 5	Exon 5	ttttttttttttttctgtttgagctcttctctggcttcttgaaatactttcctggggcacacagg	C6-Aminolink
H041012-004-E06	*CYP2B6/7*	Exon 5	Exon 5	ttttttttttttttcacagtgtggagaagcaccgtgaaaccctggaccccag	C7-Aminolink
H041012-004-F06	*CYP2B6/7*	Exon 5	Exon 5	ttttttttttttttcccccarggacctcatcgacacctacctgctccacatggaaaaa	C6-Aminolink
H041012-004-G06	*CYP2B6/7*	Exon 5	Intron 5	ttttttttttttttttacctgctccacatggaaaaagtggggtctgggagagga	C6-Aminolink
H041012-004-H06	*CYP2B6/7*	Exon 5	Exon 7	ttttttttttttttttacctacctgctccacatggaaaaaagagagtctacagggagattgaacagg	C6-Aminolink
H041012-004-A07	*CYP2B6/7*	Exon 5	Exon 6	ttttttttttttttttcctacctgctccacatggaaaaaagagaaatccaacscacacagtg	C6-Aminolink
H041012-004-B07	*CYP2B6*	Intron 5	Exon 6	ttttttttttttttttttccttccctactgtggacgcaggagaaatccaacgcacaca	C6-Aminolink
H041012-004-D07	*CYP2B6*	Exon 6	Exon 6	ttttttttttttttttaattcagccaccagaacctcaacctcaacacgctctcgctcttctttg	C6-Aminolink
H041012-004-F07	*CYP2B6/7*	Exon 6	Exon 6	ttttttttttttttttcctcawcmtcaacacgctctcgctcttctttgctggcactgaga	C6-Aminolink
H041012-004-G07	*CYP2B6/7*	Exon 6	Exon 6	ttttttttttttttttcactctccgctacggcttcctgctcatgctcaaataccctcatgtygcar	C6-Aminolink
H041012-004-H07	*CYP2B6/7*	Exon 6	Intron 6	ttttttttttttttttaataccctcatgtygcargtgggccagggacagcca	C6-Aminolink
H041012-004-A08	*CYP2B6/7*	Exon 6	Exon 8	ttttttttttttttttgctcaaataccctcatgtygcargacacagaagtatttctcawcctgag	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-004-B08	*CYP2B6/7*	Exon 6	Exon 7	ttttttttttttttttgctcaaataccctcatgtygcaragagagtctacagggagattgaacag	C6-Aminolink
H041012-004-C08	*CYP2B6/7*	Intron 6	Exon 7	ttttttttttttttttttcattggtcttctttttctgtacagagagagtctacagggagattgaacag	C6-Aminolink
H041012-004-D08	*CYP2B6/7*	Exon 7	Exon 7	ttttttttttttttttagagagtctacagggagattgaacaggtgattggcccacatcgccct	C6-Aminolink
H041012-004-E08	*CYP2B6/7*	Exon 7	Exon 7	ttttttttttttttttgagcttcatgaccgagccaaaatgccatacacagaggcagtcatctatgag	C6-Aminolink
H041012-004-F08	*CYP2B6/7*	Exon 7	Exon 7	ttttttttttttttttcattgtcacccaacacaccagcttccgagggtacatcatccccaag	C6-Aminolink
H041012-004-G08	*CYP2B6/7*	Exon 7	Intron 7	ttttttttttttttttcgagggtacatcatccccaaggtaagaccggctggaacc	C6-Aminolink
H041012-004-H08	*CYP2B6/7*	Exon 7	Exon 9	ttttttttttttttttgagggtacatcatccccaagggaagcggatttgtcttggtg	C6-Aminolink
H041012-004-A09	*CYP2B6/7*	Exon 7	Exon 8	ttttttttttttttttgcttccgagggtacatcatccccaaggacacagaagtatttctcawcctg	C6-Aminolink
H041012-004-B09	*CYP2B6/7*	Intron 7	Exon 8	ttttttttttttttttctttgatcttgtgatcctccctcaggacacagaagtatttctcawcctg	C6-Aminolink
H041012-004-C09	2B6/7	Exon 8	Exon 8	ttttttttttttttttgacacagaagtatttctcawcctgagcactgctctccatgacccacactac	C6-Aminolink
H041012-004-D09	2B6/7	Exon 8	Exon 8	ttttttttttttttttgaaaaaccagacgccttcaatcctgaccactttctggatgccaat	C6-Aminolink
H041012-004-E09	2B6/7	Exon 8	Exon 8	ttttttttttttttttatgccaatggggcactgaaaaagactgaagctttatccccttctccttag	C6-Aminolink
H041012-004-F09	2B6/7	Exon8	Intron 8	ttttttttttttttttgcttttatccccttctccttaggtaagctggaccacaatttctttcccag	C6-Aminolink
H041012-004-G09	2B6/7	Exon 8	Exon 9	ttttttttttttttttaagcttttatccccttctccttagggaagcggatttgtcttggtg	C6-Aminolink
H041012-004-H09	2B6/7	Intron 8	Exon 9	ttttttttttttttttgtgaccttctgtgtccacagggaagcggatttgtcttgg	C6-Aminolink
H041012-004-A10	2B6/7	Exon 9	Exon 9	ttttttttttttttttggaagcggatttgtcttggtgaaggcatcgcccgt	C6-Aminolink

Name	Specificity	From 3'	To 5'	Sequence 3'-5'	Modification
H041012-004-B10	2B6/7	Exon 9	Exon 9	ttttttttttttttttatctgacacccc aggagtgtggtgtgggcaaa atacccccaac	C6-Aminolink
H041012-004-C10	2B6/7	Exon 9	Exon 9	ttttttttttttttttaaaatacccccca acataccagatcygcttcctgc cccgctga	C6-Aminolink
H041012-004-D10	2B6/7	Exon 9	UTR	ttttttttttttttttgcttcctgccccg ctgaaggggctgagggaag gg	C6-Aminolink
H041012-005-F08	HybrPr			ttttttttttttttttgaacgagtcaa gcgaacgagtcaagcgaac gagtcaggag	C6-Aminolink
H041012-005-G08	NegHybrPr			ttttttttttttttttggatgttcagga accagtacgacacagatgtc aagtcgaatcagtacgacgg	C6-Aminolink
SpotPr	SpotPr-Cy3 labelled			ttttttttttttttctagacagccact cata	C6-Aminolink

7 Reference List

Alexandre,I., Hamels,S., Dufour,S., Collet,J., Zammatteo,N., De,L.F., Gala,J.L., and Remacle,J. (2001). Colorimetric silver detection of DNA microarrays. Anal. Biochem *295*, 1-8.

Ariyoshi,N., Miyazaki,M., Toide,K., Sawamura,Y., and Kamataki,T. (2001). A single nucleotide polymorphism of *CYP2B6* found in Japanese enhances catalytic activity by autoactivation. Biochem Biophys. Res Commun. *281*, 1256-1260.

Balakirev,E.S. and Ayala,F.J. (2003). Pseudogenes: are they "junk" or functional DNA? Annu. Rev Genet *37*, 123-151.

Bischof,J.M., Chiang,A.P., Scheetz,T.E., Stone,E.M., Casavant,T.L., Sheffield,V.C., and Braun,T.A. (2006). Genome-wide identification of pseudogenes capable of disease-causing gene conversion. Hum Mutat. *27*, 545-552.

Black,D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev Biochem. *72*, 291-336.

Blanchard A.P, Kaiser R.J., and Hood L.E. High-density oligonucleotide arrays. 11[6/7], 687-690. 1996. Biosens Bioelectron.
Ref Type: Generic

Blanchette,M., Green,R.E., Brenner,S.E., and Rio,D.C. (2005). Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. Genes Dev. *19*, 1306-1314.

Blencowe,B.J. (2006). Alternative splicing: new insights from global analyses. Cell *126*, 37-47.

Blievernicht,J.K., Schaeffeler,E., Klein,K., Eichelbaum,M., Schwab,M., and Zanger,U.M. (2007). MALDI-TOF mass spectrometry for multiplex genotyping of *CYP2B6* single-nucleotide polymorphisms. Clin Chem *53*, 24-33.

Bodrossy,L. (2003). Diagnostic oligonucleotide microarrays for microbiology. In E. Blalock (ed.), A Beginner's Guide to Microarrays, Kluwer Academic Publishers, New York), pp. 43-92.

Bradford,L.D. (2002). *CYP2D6* allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics *3*, 229-243.

Brett,D., Hanke,J., Lehmann,G., Haase,S., Delbruck,S., Krueger,S., Reich,J., and Bork,P. (2000). EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett. *474*, 83-86.

Brinkman,B.M. (2004). Splice variants as cancer biomarkers. Clin Biochem. *37*, 584-594.

Cartegni,L., Chew,S.L., and Krainer,A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Cartegni,L., Wang,J., Zhu,Z., Zhang,M.Q., and Krainer,A.R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res *31*, 3568-3571.

Castle,J., Garrett-Engele,P., Armour,C.D., Duenwald,S.J., Loerch,P.M., Meyer,M.R., Schadt,E.E., Stoughton,R., Parrish,M.L., Shoemaker,D.D., and Johnson,J.M. (2003). Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. Genome Biol *4*, R66.

C.elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012-2018.

Chang,T.K., Weber,G.F., Crespi,C.L., and Waxman,D.J. (1993). Differential activation of cyclophosphamide and ifosphamide by cytochromes P-450 2B and 3A in human liver microsomes. Cancer Res *53*, 5629-5637.

Cheung,V.G., Morley,M., Aguilar,F., Massimi,A., Kucherlapati,R., and Childs,G. (1999). Making and reading microarrays. Nat Genet *21*, 15-19.

Chiara,M.D., Palandjian,L., Feld,K.R., and Reed,R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. EMBO J *16*, 4746-4759.

Chua,K. and Reed,R. (2001). An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. Mol Cell Biol *21*, 1509-1514.

Clark,T.A., Sugnet,C.W., and Ares,M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science *296*, 907-910.

Colgan,D.F. and Manley,J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev. *11*, 2755-2766.

Coon,M.J., Ding,X.X., Pernecky,S.J., and Vaz,A.D. (1992). Cytochrome P450: progress and predictions. FASEB J *6*, 669-673.

Court MH, Duan,S.X., Hesse,L.M., Venkatakrishnan,K., and Greenblatt,D.J. (2001). Cytochrome P-450 2B6 is responsible for interindividual variability of propofol hydroxylation by human liver microsomes. Anesthesiology *94*, 110-119.

Cuperlovic-Culf,M., Belacel,N., Culf,A.S., and Ouellette,R.J. (2006). Microarray analysis of alternative splicing. OMICS. *10*, 344-357.

Desta,Z., Saussele,T., Ward,B., Blievernicht,J., Li,L., Klein,K., Flockhart,D.A., and Zanger,U.M. (2007). Impact of CYP2B6 polymorphism on hepatic efavirenz metabolism in vitro. Pharmacogenomics *8*, 547-558.

Deutsch,M. and Long,M. (1999). Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res *27*, 3219-3228.

Eichelbaum,M., Spannbrucker,N., and Dengler,H.J. (1975). Proceedings: N-oxidation of sparteine in man and its interindividual differences. Naunyn Schmiedebergs Arch Pharmacol *287 Suppl*, R94.

Eichelbaum,M., Spannbrucker,N., Steincke,B., and Dengler,H.J. (1979). Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. Eur J Clin Pharmacol *16*, 183-187.

Endrizzi,K., Fischer,J., Klein,K., Schwab,M., Nussler,A., Neuhaus,P., Eichelbaum,M., and Zanger,U.M. (2002). Discriminative quantification of cytochrome P450*2D6* and *2D7/8* pseudogene expression by TaqMan real-time reverse transcriptase polymerase chain reaction. Anal. Biochem. *300*, 121-131.

Erickson,D.A., Mather,G., Trager,W.F., Levy,R.H., and Keirns,J.J. (1999). Characterization of the in vitro biotransformation of the HIV-1 reverse transcriptase inhibitor nevirapine by human hepatic cytochromes P-450. Drug Metab Dispos. *27*, 1488-1495.

Esteller,M. (2007). Epigenetic gene silencing in cancer: the DNA hypermethylome. Hum Mol Genet *16 Spec No 1*, R50-R59.

Evert,B., Eichelbaum,M., Haubruck,H., and Zanger,U.M. (1997). Functional properties of CYP2D6 1 (wild-type) and CYP2D6 7 (His324Pro) expressed by recombinant baculovirus in insect cells. Naunyn Schmiedebergs Arch Pharmacol *355*, 309-318.

Faucette,S.R., Sueyoshi,T., Smith,C.M., Negishi,M., LeCluyse,E.L., and Wang,H. (2006). Differential regulation of hepatic *CYP2B6* and *CYP3A4* genes by constitutive androstane receptor but not pregnane X receptor. J Pharmacol Exp Ther *317*, 1200-1209.

Faucette,S.R., Zhang,T.C., Moore,R., Sueyoshi,T., Omiecinski,C.J., LeCluyse,E.L., Negishi,M., and Wang,H. (2007). Relative activation of human pregnane X receptor versus constitutive androstane receptor defines distinct classes of CYP2B6 and CYP3A4 inducers. J Pharmacol Exp Ther *320*, 72-80.

Faustino,N.A. and Cooper,T.A. (2003). Pre-mRNA splicing and human disease. Genes Dev. *17*, 419-437.

Fehlbaum,P., Guihal,C., Bracco,L., and Cochet,O. (2005). A microarray configuration to quantify expression levels and relative abundance of splice variants. Nucleic Acids Res *33*, e47.

Gaedigk,A., Bradford,L.D., Marcucci,K.A., and Leeder,J.S. (2002). Unique CYP2D6 activity distribution and genotype-phenotype discordance in black Americans. Clin Pharmacol Ther *72*, 76-89.

Gaedigk,A., Gaedigk,R., and Leeder,J.S. (2005). *CYP2D7* splice variants in human liver and brain: does *CYP2D7* encode functional protein? Biochem. Biophys. Res Commun. *336*, 1241-1250.

Gardina,P.J., Clark,T.A., Shimada,B., Staples,M.K., Yang,Q., Veitch,J., Schweitzer,A., Awad,T., Sugnet,C., Dee,S., Davies,C., Williams,A., and Turpaz,Y.

(2006). Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics *7*, 325.

Gatanaga,H., Hayashida,T., Tsuchiya,K., Yoshino,M., Kuwahara,T., Tsukada,H., Fujimoto,K., Sato,I., Ueda,M., Horiba,M., Hamaguchi,M., Yamamoto,M., Takata,N., Kimura,A., Koike,T., Gejyo,F., Matsushita,S., Shirasaka,T., Kimura,S., and Oka,S. (2007). Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 *2B6*6* and **26*. Clin Infect Dis *45*, 1230-1237.

Gerhard,D.S., Wagner,L., Feingold,E.A., Shenmen,C.M., Grouse,L.H., Schuler,G., *et al.* and Malek,J. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Res *14*, 2121-2127.

Gervot,L., Rochat,B., Gautier,J.C., Bohnenstengel,F., Kroemer,H., de,B., V, Martin,H., Beaune,P., and de,W., I (1999). Human *CYP2B6*: expression, inducibility and catalytic activities. Pharmacogenetics *9*, 295-306.

Giacomoni,D. and Spiegelman,S. (1962). Origin and biologic individuality of the genetic dictionary. Science *138*, 1328-1331.

Gonzalez FJ, Crespi CL, Czerwinski M, Gelboin HV. (1992). Analysis of human cytochrome P450 catalytic activities and expression. Tohoku J Exp Med. *168*, 67-72.

Gonzalez,F.J., Skoda,R.C., Kimura,S., Umeno,M., Zanger,U.M., Nebert,D.W., Gelboin,H.V., Hardwick,J.P., and Meyer,U.A. (1988). Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. Nature *331*, 442-446.

Graveley,B.R. (2000). Sorting out the complexity of SR protein functions. RNA. *6*, 1197-1211.

Graveley,B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. Trends Genet *17*, 100-107.

Gregory,S.G., Barlow,K.F., McLay,K.E., Kaul,R., Swarbreck,D., Dunham,A., *et al.* and Prigmore,E. (2006). The DNA sequence and biological annotation of human chromosome 1. Nature *441*, 315-321.

Grimm,V., Ezaki,S., Susa,M., Knabbe,C., Schmid,R.D., and Bachmann,T.T. (2004). Use of DNA microarrays for rapid genotyping of *TEM* beta-lactamases that confer resistance. J. Clin. Microbiol. *42*, 3766-3774.

Haas,D.W. (2006). Human genetic variability and HIV treatment response. Curr HIV. /AIDS Rep. *3*, 53-58.

Haas,D.W., Smeaton,L.M., Shafer,R.W., Robbins,G.K., Morse,G.D., Labbe,L., Wilkinson,G.R., Clifford,D.B., D'Aquila,R.T., De,G., V, Pollard,R.B., Merigan,T.C., Hirsch,M.S., George,A.L., Jr., Donahue,J.P., and Kim,R.B. (2005). Pharmacogenetics of long-term responses to antiretroviral regimens containing Efavirenz and/or Nelfinavir: an Adult Aids Clinical Trials Group Study. J Infect Dis *192*, 1931-1942.

Hattori,M. (2005). [Finishing the euchromatic sequence of the human genome]. Tanpakushitsu Kakusan Koso *50*, 162-168.

He,H. and Lehming,N. (2003). Global effects of histone modifications. Brief Funct Genomic Proteomic *2*, 234-243.

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

Heller,M.J. (2002). DNA microarray technology: devices, systems, and applications. Annu. Rev Biomed. Eng *4*, 129-153.

Hesse,L.M., He,P., Krishnaswamy,S., Hao,Q., Hogan,K., von Moltke,L.L., Greenblatt,D.J., and Court MH (2004). Pharmacogenetic determinants of interindividual variability in bupropion hydroxylation by cytochrome P450 2B6 in human liver microsomes. Pharmacogenetics *14*, 225-238.

Hlavica,P. (2004). Models and mechanisms of O-O bond activation by cytochrome P450. A critical assessment of the potential role of multiple active intermediates in oxidative catalysis. Eur J Biochem *271*, 4335-4360.

Hodgkin,J. (2001). What does a worm want with 20000 genes? Genome Biol *2*, COMMENT2008.

Hodgson,E. and Rose,R.L. (2007). The importance of cytochrome P450 2B6 in the human metabolism of environmental chemicals. Pharmacol Ther *113*, 420-428.

Hoffman,S.M., Nelson,D.R., and Keeney,D.S. (2001). Organization, structure and evolution of the *CYP2* gene cluster on human chromosome 19. Pharmacogenetics *11*, 687-698.

Hofmann,M.H., Blievernicht,J.K., Klein,K., Saussele,T., Schaeffeler,E., Schwab,M., and Zanger,U.M. (2008). Aberrant splicing caused by single nucleotide polymorphism c.516G>T [Q172H], a marker of *CYP2B6*6*, is responsible for decreased expression and activity of CYP2B6 in Liver. J. Pharmacol. Exp. Ther. 325, 284-292

Hu,G.K., Madore,S.J., Moldover,B., Jatkoe,T., Balaban,D., Thomas,J., and Wang,Y. (2001). Predicting splice variant from DNA chip expression data. Genome Res *11*, 1237-1245.

Hustert,E., Haberl,M., Burk,O., Wolbold,R., He,Y.Q., Klein,K., Nuessler,A.C., Neuhaus,P., Klattig,J., Eiselt,R., Koch,I., Zibat,A., Brockmoller,J., Halpert,J.R., Zanger,U.M., and Wojnowski,L. (2001). The genetic determinants of the *CYP3A5* polymorphism. Pharmacogenetics *11*, 773-779.

Ingelman-Sundberg,M. (1999). Duplication, multiduplication, and amplification of genes encoding drug-metabolizing enzymes: evolutionary, toxicological, and clinical pharmacological aspects. Drug Metab Rev *31*, 449-459.

Ingelman-Sundberg,M. (2004). Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. Trends Pharmacol Sci. *25*, 193-200.

Ingelman-Sundberg,M. (2005). Genetic polymorphisms of cytochrome P450 *2D6* (*CYP2D6*): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J *5*, 6-13.

Ingelman-Sundberg,M. and Rodriguez-Antona,C. (2005). Pharmacogenetics of drug-metabolizing enzymes: implications for a safer and more effective drug therapy. Philos. Trans. R. Soc Lond B Biol Sci. *360*, 1563-1570.

Jain,K.K. (2000). Applications of biochip and microarray systems in pharmacogenomics. Pharmacogenomics *1*, 289-307.

Jinno,H., Tanaka-Kagawa,T., Ohno,A., Makino,Y., Matsushima,E., Hanioka,N., and Ando,M. (2003). Functional characterization of cytochrome P450 *2B6* allelic variants. Drug Metab Dispos. *31*, 398-403.

Johansson,I., Lundqvist,E., Bertilsson,L., Dahl,M.L., Sjoqvist,F., and Ingelman-Sundberg,M. (1993). Inherited amplification of an active gene in the cytochrome P450 *CYP2D* locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad. Sci. U S. A *90*, 11825-11829.

Johnson,J.M., Castle,J., Garrett-Engele,P., Kan,Z., Loerch,P.M., Armour,C.D., Santos,R., Schadt,E.E., Stoughton,R., and Shoemaker,D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science *302*, 2141-2144.

Kagimoto,M., Heim,M., Kagimoto,K., Zeugin,T., and Meyer,U.A. (1990). 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 *265*, 17209-17214.

Kampa,D., Cheng,J., Kapranov,P., Yamanaka,M., Brubaker,S., Cawley,S., Drenkow,J., Piccolboni,A., Bekiranov,S., Helt,G., Tammana,H., and Gingeras,T.R. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res *14*, 331-342.

Kan,Z., Garrett-Engele,P.W., Johnson,J.M., and Castle,J.C. (2005). Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. Nucleic Acids Res *33*, 5659-5666.

Kapranov,P., Cawley,S.E., Drenkow,J., Bekiranov,S., Strausberg,R.L., Fodor,S.P., and Gingeras,T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science *296*, 916-919.

Kharasch,E.D., Hoffer,C., Whittington,D., and Sheffels,P. (2004). Role of hepatic and intestinal cytochrome P450 3A and 2B6 in the metabolism, disposition, and miotic effects of methadone. Clin Pharmacol Ther *76*, 250-269.

Khelifi,A., Duret,L., and Mouchiroud,D. (2005). Hoppsigen: a database of human and mouse processed pseudogenes. Nucleic Acids Res *33*, D59-D66.

Kim,H., Klein,R., Majewski,J., and Ott,J. (2004). Estimating rates of alternative splicing in mammals and invertebrates. Nat Genet *36*, 915-916.

Kimura,S., Umeno,M., Skoda,R.C., Meyer,U.A., and Gonzalez,F.J. (1989). 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 *45*, 889-904.

Kirchheiner,J., Klein,C., Meineke,I., Sasse,J., Zanger,U.M., Murdter,T.E., Roots,I., and Brockmoller,J. (2003). Bupropion and 4-OH-bupropion pharmacokinetics in relation to genetic polymorphisms in *CYP2B6*. Pharmacogenetics *13*, 619-626.

Kirschbaum-Slager,N., Lopes,G.M., Galante,P.A., Riggins,G.J., and de Souza,S.J. (2004). Splicing factors are differentially expressed in tumors. Genet Mol Res *3*, 512-520.

Kirschbaum-Slager,N., Parmigiani,R.B., Camargo,A.A., and de Souza,S.J. (2005). Identification of human exons overexpressed in tumors through the use of genome and expressed sequence data. Physiol Genomics *21*, 423-432.

Klein,K., Lang,T., Saussele,T., Barbosa-Sicard,E., Schunck,W.H., Eichelbaum,M., Schwab,M., and Zanger,U.M. (2005). Genetic variability of *CYP2B6* in populations of African and Asian origin: allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. Pharmacogenet. Genomics *15*, 861-873.

Klingenberg,M. (1958). Pigments of rat liver microsomes. Arch Biochem Biophys. *75*, 376-386.

Koch,W.H. (2004). Technology platforms for pharmacogenomic diagnostic assays. Nat Rev Drug Discov. *3*, 749-761.

Krawczak,M., Reiss,J., and Cooper,D.N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet *90*, 41-54.

Kuehl,P., Zhang,J., Lin,Y., Lamba,J., Assem,M., Schuetz,J., Watkins,P.B., Daly,A., Wrighton,S.A., Hall,S.D., Maurel,P., Relling,M., Brimer,C., Yasuda,K., Venkataramanan,R., Strom,S., Thummel,K., Boguski,M.S., and Schuetz,E. (2001). Sequence diversity in *CYP3A* promoters and characterization of the genetic basis of polymorphic *CYP3A5* expression. Nat Genet *27*, 383-391.

Lamba,V., Lamba,J., Yasuda,K., Strom,S., Davila,J., Hancock,M.L., Fackenthal,J.D., Rogan,P.K., Ring,B., Wrighton,S.A., and Schuetz,E.G. (2003). Hepatic *CYP2B6* expression: gender and ethnic differences and relationship to *CYP2B6* genotype and *CAR* (constitutive androstane receptor) expression. J Pharmacol Exp Ther *307*, 906-922.

Lang,T., Klein,K., Fischer,J., Nussler,A.K., Neuhaus,P., Hofmann,U., Eichelbaum,M., Schwab,M., and Zanger,U.M. (2001). Extensive genetic polymorphism in the human *CYP2B6* gene with impact on expression and function in human liver. Pharmacogenetics *11*, 399-415.

Lang,T., Klein,K., Richter,T., Zibat,A., Kerb,R., Eichelbaum,M., Schwab,M., and Zanger,U.M. (2004). Multiple novel nonsynonymous *CYP2B6* gene polymorphisms in Caucasians: demonstration of phenotypic null alleles. J Pharmacol Exp Ther *311*, 34-43.

Le,K., Mitsouras,K., Roy,M., Wang,Q., Xu,Q., Nelson,S.F., and Lee,C. (2004). Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. Nucleic Acids Res *32*, e180.

Lee,C. and Roy,M. (2004). Analysis of alternative splicing with microarrays: successes and challenges. Genome Biol *5*, 231.

Lewis,D.F. (2004). 57 varieties: the human cytochromes P450. Pharmacogenomics. *5*, 305-318.

Lewis,J.D., Gunderson,S.I., and Mattaj,I.W. (1995). The influence of 5' and 3' end structures on pre-mRNA metabolism. J Cell Sci. Suppl *19*, 13-19.

Li,C. and Wong,W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad. Sci. U S. A *98*, 31-36.

Lipshutz,R.J., Fodor,S.P., Gingeras,T.R., and Lockhart,D.J. (1999). High density synthetic oligonucleotide arrays. Nat Genet *21*, 20-24.

Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H., and Brown,E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*, 1675-1680.

Maas,S., Kawahara,Y., Tamburro,K.M., and Nishikura,K. (2006). A-to-I RNA editing and human disease. RNA. Biol *3*, 1-9.

Mahgoub,A., Hirsch,P.F., and Munson,P.L. (1997). Calcium-lowering action of glucocorticoids in adrenalectomized-parathyroidectomized rats. Specificity and relative potency of natural and synthetic glucocorticoids. Endocrine. *6*, 279-283.

Maniatis,T. and Tasic,B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature *418*, 236-243.

Marshall,E. (2003). Preventing toxicity with a gene test. Science *302*, 588-590.

Marzolini,C., Telenti,A., Decosterd,L.A., Greub,G., Biollaz,J., and Buclin,T. (2001). Efavirenz plasma levels can predict treatment failure and central nervous system side effects in HIV-1-infected patients. AIDS *15*, 71-75.

Meyer,U.A. (2000). Pharmacogenetics and adverse drug reactions. Lancet *356*, 1667-1671.

Michaud,S. and Reed,R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev. *5*, 2534-2546.

Miles,J.S., McLaren,A.W., and Wolf,C.R. (1989). Alternative splicing in the human cytochrome *P450IIB6* gene generates a high level of aberrant messages. Nucleic Acids Res *17*, 8241-8255.

Miremadi,A., Oestergaard,M.Z., Pharoah,P.D., and Caldas,C. (2007). Cancer genetics of epigenetic genes. Hum Mol Genet *16 Spec No 1*, R28-R49.

Mironov,A.A., Fickett,J.W., and Gelfand,M.S. (1999). Frequent alternative splicing of human genes. Genome Res *9*, 1288-1293.

Mizutani,T. (2003). PM frequencies of major *CYPs* in Asians and Caucasians. Drug Metab Rev *35*, 99-106.

Mockler,T.C. and Ecker,J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. Genomics *85*, 1-15.

Modrek,B. and Lee,C. (2002). A genomic view of alternative splicing. Nat Genet *30*, 13-19.

Modrek,B., Resch,A., Grasso,C., and Lee,C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res *29*, 2850-2859.

Myers,E.W., Sutton,G.G., Delcher,A.L., Dew,I.M., Fasulo,D.P., Flanigan,M.J., Kravitz,S.A., Mobarry,C.M., Reinert,K.H., Remington,K.A., Anson,E.L., Bolanos,R.A., Chou,H.H., Jordan,C.M., Halpern,A.L., Lonardi,S., Beasley,E.M., Brandon,R.C., Chen,L., Dunn,P.J., Lai,Z., Liang,Y., Nusskern,D.R., Zhan,M., Zhang,Q., Zheng,X., Rubin,G.M., Adams,M.D., and Venter,J.C. (2000). A whole-genome assembly of Drosophila. Science *287*, 2196-2204.

Nagao,K., Togawa,N., Fujii,K., Uchikawa,H., Kohno,Y., Yamada,M., and Miyashita,T. (2005). Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the *PTCH* gene with exon junction microarrays. Hum Mol. Genet *14*, 3379-3388.

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

Newcombe,R.G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med *17*, 857-872.

Novoyatleva,T., Tang,Y., Rafalska,I., and Stamm,S. (2006). Pre-mRNA missplicing as a cause of human disease. Prog. Mol Subcell. Biol *44*, 27-46.

Owen,A., Pirmohamed,M., Khoo,S.H., and Back,D.J. (2006). Pharmacogenetics of HIV therapy. Pharmacogenet. Genomics *16*, 693-703.

Pan,Q., Shai,O., Misquitta,C., Zhang,W., Saltzman,A.L., Mohammad,N., Babak,T., Siu,H., Hughes,T.R., Morris,Q.D., Frey,B.J., and Blencowe,B.J. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol. Cell *16*, 929-941.

Pease,A.C., Solas,D., Sullivan,E.J., Cronin,M.T., Holmes,C.P., and Fodor,S.P. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. U. S. A *91*, 5022-5026.

Quelle,D.E., Zindy,F., Ashmun,R.A., and Sherr,C.J. (1995). Alternative reading frames of the *INK4a* tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. Cell *83*, 993-1000.

Query,C.C., Strobel,S.A., and Sharp,P.A. (1996). Three recognition events at the branch-site adenine. EMBO J *15*, 1392-1402.

Rae,J.M., Soukhova,N.V., Flockhart,D.A., and Desta,Z. (2002). Triethylenethiophosphoramide is a specific inhibitor of cytochrome P450 2B6: implications for cyclophosphamide metabolism. Drug Metab Dispos. *30*, 525-530.

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

Relogio,A., Ben-Dov,C., Baum,M., Ruggiu,M., Gemund,C., Benes,V., Darnell,R.B., and Valcarcel,J. (2005). Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. J Biol Chem. *280*, 4779-4784.

Richter,T., Murdter,T.E., Heinkele,G., Pleiss,J., Tatzel,S., Schwab,M., Eichelbaum,M., and Zanger,U.M. (2004). Potent mechanism-based inhibition of human CYP2B6 by clopidogrel and ticlopidine. J Pharmacol Exp Ther *308*, 189-197.

Richter,T., Schwab,M., Eichelbaum,M., and Zanger,U.M. (2005). Inhibition of human CYP2B6 by N,N',N''-triethylenethiophosphoramide is irreversible and mechanism-based. Biochem Pharmacol *69*, 517-524.

Ristaldi,M.S., Pirastu,M., Rosatelli,C., Monni,G., Erlich,H., Saiki,R., and Cao,A. (1989). Prenatal diagnosis of beta-thalassaemia in Mediterranean populations by dot blot analysis with DNA amplification and allele specific oligonucleotide probes. Prenat. Diagn. *9*, 629-638.

Rodriguez-Novoa,S., Barreiro,P., Jimenez-Nacher,I., Rendon,A., and Soriano,V. (2005). Pharmacogenetics in HIV therapy. AIDS Rev *7*, 103-112.

Rotger,M., Csajka,C., and Telenti,A. (2006). Genetic, ethnic, and gender differences in the pharmacokinetics of antiretroviral agents. Curr HIV. /AIDS Rep. *3*, 118-125.

Rotger,M., Tegude,H., Colombo,S., Cavassini,M., Furrer,H., Decosterd,L., Blievernicht,J., Saussele,T., Gunthard,H.F., Schwab,M., Eichelbaum,M., Telenti,A., and Zanger,U.M. (2007). Predictive value of known and novel alleles of *CYP2B6* for efavirenz plasma concentrations in HIV-infected individuals. Clin Pharmacol Ther *81*, 557-566.

Roy,P., Yu,L.J., Crespi,C.L., and Waxman,D.J. (1999). Development of a substrate-activity based approach to identify the major human liver P-450 catalysts of cyclophosphamide and ifosfamide activation based on cDNA-expressed activities and liver microsomal P-450 profiles. Drug Metab Dispos. *27*, 655-666.

Saussele,T., Burk,O., Blievernicht,J.K., Klein,K., Nussler,A., Nussler,N., Hengstler,J.G., Eichelbaum,M., Schwab,M., and Zanger,U.M. (2007). Selective induction of human hepatic cytochromes P450 *2B6* and *3A4* by metamizole. Clin Pharmacol Ther *82*, 265-274.

Schena,M., Shalon,D., Davis,R.W., and Brown,P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Sharp,P.A. (1994). Split genes and RNA splicing. Cell *77*, 805-815.

Shchepinov,M.S., Case-Green,S.C., and Southern,E.M. (1997). Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. Nucleic Acids Res *25*, 1155-1161.

Shoemaker,D.D., Schadt,E.E., Armour,C.D., He,Y.D., Garrett-Engele,P., McDonagh,P.D., Loerch,P.M., Leonardson,A., Lum,P.Y., Cavet,G., Wu,L.F., Altschuler,S.J., Edwards,S., King,J., Tsang,J.S., Schimmack,G., Schelter,J.M., Koch,J., Ziman,M., Marton,M.J., Li,B., Cundiff,P., Ward,T., Castle,J., Krolewski,M., Meyer,M.R., Mao,M., Burchard,J., Kidd,M.J., Dai,H., Phillips,J.W., Linsley,P.S., Stoughton,R., Scherer,S., and Boguski,M.S. (2001). Experimental annotation of the human genome using microarray technology. Nature *409*, 922-927.

Sim,S.C. and Ingelman-Sundberg,M. (2006). The human cytochrome P450 Allele Nomenclature Committee Web site: submission criteria, procedures, and objectives. Methods Mol Biol *320*, 183-191.

Smith,C.M., Faucette,S.R., Wang,H., and LeCluyse,E.L. (2005). Modulation of UDP-glucuronosyltransferase 1A1 in primary human hepatocytes by prototypical inducers. J Biochem Mol Toxicol *19*, 96-108.

Smith,P.J., Zhang,C., Wang,J., Chew,S.L., Zhang,M.Q., and Krainer,A.R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet *15*, 2490-2508.

Southern,E., Mir,K., and Shchepinov,M. (1999). Molecular interactions on microarrays. Nat. Genet. *21*, 5-9.

Southern,E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503-517.

Southern,E.M. (2001). DNA microarrays. History and overview. Methods Mol Biol *170*, 1-15.

Southern,E.M. and Maskos,U. (1994). Parallel synthesis and analysis of large numbers of related chemical compounds: applications to oligonucleotides. J. Biotechnol. *35*, 217-227.

Srebrow,A. and Kornblihtt,A.R. (2006). The connection between splicing and cancer. J Cell Sci. *119*, 2635-2641.

Srinivasan,K., Shiue,L., Hayes,J.D., Centers,R., Fitzwater,S., Loewen,R., Edmondson,L.R., Bryant,J., Smith,M., Rommelfanger,C., Welch,V., Clark,T.A., Sugnet,C.W., Howe,K.J., Mandel-Gutfreund,Y., and Ares,M., Jr. (2005). Detection and measurement of alternative splicing using splicing-sensitive microarrays. Methods *37*, 345-359.

Stears,R.L., Martinsky,T., and Schena,M. (2003). Trends in microarray analysis. Nat Med *9*, 140-145.

Stephens,M. and Donnelly,P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet *73*, 1162-1169.

Stephens,M., Smith,N.J., and Donnelly,P. (2001). A new statistical method for haplotype reconstruction from population data. Am J Hum Genet *68*, 978-989.

Stolc,V., Gauhar,Z., Mason,C., Halasz,G., van Batenburg,M.F., Rifkin,S.A., Hua,S., Herreman,T., Tongprasit,W., Barbano,P.E., Bussemaker,H.J., and White,K.P. (2004). A gene expression map for the euchromatic genome of Drosophila melanogaster. Science *306*, 655-660.

Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., *et al.* and Marra,M.A. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl. Acad. Sci. U S. A *99*, 16899-16903.

Svensson,U.S. and Ashton,M. (1999). Identification of the human cytochrome P450 enzymes involved in the in vitro metabolism of artemisinin. Br. J Clin Pharmacol *48*, 528-535.

Telenti,A. and Zanger,U.M. (2007). Pharmacogenetics of Anti-HIV Drugs. Annu. Rev Pharmacol Toxicol.

Tompkins,L.M. and Wallace,A.D. (2007). Mechanisms of cytochrome P450 induction. J Biochem Mol Toxicol *21*, 176-181.

Toscano,C., Klein,K., Blievernicht,J., Schaeffeler,E., Saussele,T., Raimundo,S., Eichelbaum,M., Schwab,M., and Zanger,U.M. (2006). Impaired expression of *CYP2D6* in intermediate metabolizers carrying the **41* allele caused by the intronic SNP 2988G>A: evidence for modulation of splicing events. Pharmacogenet. Genomics *16*, 755-766.

Tsuchiya,K., Gatanaga,H., Tachikawa,N., Teruya,K., Kikuchi,Y., Yoshino,M., Kuwahara,T., Shirasaka,T., Kimura,S., and Oka,S. (2004). Homozygous *CYP2B6 *6* (Q172H and K262R) correlates with high plasma efavirenz concentrations in HIV-1 patients treated with standard efavirenz-containing regimens. Biochem Biophys. Res Commun. *319*, 1322-1326.

Turpeinen,M., Raunio,H., and Pelkonen,O. (2006). The functional role of CYP2B6 in human drug metabolism: substrates and inhibitors in vitro, in vivo and in silico. Curr Drug Metab *7*, 705-714.

Turunen,M.M., Dunlop,T.W., Carlberg,C., and Vaisanen,S. (2007). Selective use of multiple vitamin D response elements underlies the 1 alpha,25-dihydroxyvitamin D3-mediated negative regulation of the human *CYP27B1* gene. Nucleic Acids Res *35*, 2734-2747.

Ule,J., Ule,A., Spencer,J., Williams,A., Hu,J.S., Cline,M., Wang,H., Clark,T., Fraser,C., Ruggiu,M., Zeeberg,B.R., Kane,D., Weinstein,J.N., Blume,J., and

Darnell,R.B. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat Genet *37*, 844-852.

Van Gelder,R.N., von Zastrow,M.E., Yool,A., Dement,W.C., Barchas,J.D., and Eberwine,J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad. Sci. U S. A *87*, 1663-1667.

Waddington,C.H. (1959). Canalization of development and genetic assimilation of acquired characters. Nature *183*, 1654-1655.

Wahle,E. and Ruegsegger,U. (1999). 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol Rev *23*, 277-295.

Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R., Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L., Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E., Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J., Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M., and Lander,E.S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science *280*, 1077-1082.

Wang,H., Donley,K.M., Keeney,D.S., and Hoffman,S.M. (2003a). Organization and evolution of the *Cyp2* gene cluster on mouse chromosome 7, and comparison with the syntenic human cluster. Environ. Health Perspect. *111*, 1835-1842.

Wang,H., Faucette,S., Moore,R., Sueyoshi,T., Negishi,M., and LeCluyse,E. (2004). Human constitutive androstane receptor mediates induction of *CYP2B6* gene expression by phenytoin. J Biol Chem *279*, 29295-29301.

Wang,H., Hubbell,E., Hu,J.S., Mei,G., Cline,M., Lu,G., Clark,T., Siani-Rose,M.A., Ares,M., Kulp,D.C., and Haussler,D. (2003b). Gene structure-based splice variant deconvolution using a microarray platform. Bioinformatics. *19 Suppl 1*, i315-i322.

Wang,H. and Negishi,M. (2003). Transcriptional regulation of cytochrome *p450 2B* genes by nuclear receptors. Curr Drug Metab *4*, 515-525.

Ward,B.A., Gorski,J.C., Jones,D.R., Hall,S.D., Flockhart,D.A., and Desta,Z. (2003). The cytochrome P450 2B6 (CYP2B6) is the main catalyst of efavirenz primary and secondary metabolism: implication for HIV/AIDS therapy and utility of efavirenz as a substrate marker of CYP2B6 catalytic activity. J Pharmacol Exp Ther *306*, 287-300.

Watahiki,A., Waki,K., Hayatsu,N., Shiraki,T., Kondo,S., Nakamura,M., Sasaki,D., Arakawa,T., Kawai,J., Harbers,M., Hayashizaki,Y., and Carninci,P. (2004). Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas. Nat Methods *1*, 233-239.

Weinshilboum,R. and Wang,L. (2004). Pharmacogenomics: bench to bedside. Nat Rev Drug Discov. *3*, 739-748.

Wood,A.J. and Oakey,R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. PLoS. Genet *2*, e147.

Wu,D.Y., Nozari,G., Schold,M., Conner,B.J., and Wallace,R.B. (1989). Direct analysis of single nucleotide variation in human DNA and RNA using in situ dot hybridization. DNA *8*, 135-142.

Xie,H.J., Yasar,U., Lundgren,S., Griskevicius,L., Terelius,Y., Hassan,M., and Rane,A. (2003). Role of polymorphic human CYP2B6 in cyclophosphamide bioactivation. Pharmacogenomics J *3*, 53-61.

Yamano,S., Nhamburo,P.T., Aoyama,T., Meyer,U.A., Inaba,T., Kalow,W., Gelboin,H.V., McBride,O.W., and Gonzalez,F.J. (1989). cDNA cloning and sequence and cDNA-directed expression of human P450 *IIB1*: identification of a normal and two variant cDNAs derived from the *CYP2B* locus on chromosome 19 and differential expression of the *IIB* mRNAs in human liver. Biochemistry *28*, 7340-7348.

Yanofsky,C., Carlton,B.C., Guest,J.R., Helinski,D.R., and Henning,U. (1964). On the colinearity of gene structure and protein structure. Proc Natl Acad. Sci. U S. A *51*, 266-272.

Yao,A., Charlab,R., and Li,P. (2006). Systematic identification of pseudogenes through whole genome expression evidence profiling. Nucleic Acids Res.

Yasui,H., Hayashi,S., and Sakurai,H. (2005). Possible involvement of singlet oxygen species as multiple oxidants in p450 catalytic reactions. Drug Metab Pharmacokinet *20*, 1-13.

Yeakley,J.M., Fan,J.B., Doucet,D., Luo,L., Wickham,E., Ye,Z., Chee,M.S., and Fu,X.D. (2002). Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol. *20*, 353-358.

Yue,H., Eastman,P.S., Wang,B.B., Minor,J., Doctolero,M.H., Nuttall,R.L., Stack,R., Becker,J.W., Montgomery,J.R., Vainer,M., and Johnston,R. (2001). An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. Nucleic Acids Res *29*, E41.

Zanger,U.M., and Hofmann,M.H. (2008). Polymorphic Cytochromes P450 *CYP2B6* and *CYP2D6*: Recent Advances on Single Nucleotide Polymorphisms Affecting Splicing. Acta Chim. Slov., 55, 38-44

Zanger,U.M., Klein,K., Saussele,T., Blievernicht,J., Hofmann,M.H., and Schwab,M. (2007). Polymorphic *CYP2B6*: molecular mechanisms and emerging clinical significance. Pharmacogenomics *8*, 743-759.

Zanger,U.M., Raimundo,S., and Eichelbaum,M. (2004). Cytochrome P450 *2D6*: overview and update on pharmacology, genetics, biochemistry. Naunyn Schmiedebergs Arch. Pharmacol *369*, 23-37.

Zhang,Z., Carriero,N., Zheng,D., Karro,J., Harrison,P.M., and Gerstein,M. (2006). PseudoPipe: an automated pseudogene identification pipeline. Bioinformatics. *22*, 1437-1439.

Zhang,Z. and Gerstein,M. (2004). Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet Dev. *14*, 328-335.

Zhang,Z., Harrison,P.M., Liu,Y., and Gerstein,M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res *13*, 2541-2558.

Zukunft,J., Lang,T., Richter,T., Hirsch-Ernst,K.I., Nussler,A.K., Klein,K., Schwab,M., Eichelbaum,M., and Zanger,U.M. (2005). A natural *CYP2B6* TATA box polymorphism (-82T--> C) leading to enhanced transcription and relocation of the transcriptional start site. Mol Pharmacol *67*, 1772-1782.

8 Acknowledgement - Danksagung

Die vorliegende Dissertation wurde vom August 2004 bis Dezember 2007 am Dr. Margarete Fischer-Bosch Institut für Klinische Pharmakologie (IKP) in Stuttgart, Universität Tübingen und in Kooperation mit dem Institut für Technische Biochemie, an der Universität Stuttgart angefertigt. An dieser Stelle bedanke ich mich bei allen Mitarbeitern der zwei Institute für die gute Zusammenarbeit und freundschaftliche Atmosphäre.

Ein besonderer Dank gilt dem Leiter des Dr. Margarete Fischer-Bosch Instituts für Klinische Pharmakologie Herrn Prof. Matthias Schwab sowie seinem Vorgänger Herrn Prof. Michael Eichelbaum für die Möglichkeit, meine Doktorarbeit an ihrem Institut anfertigen zu können. Herrn Prof. Schwab danke ich sehr für sein stetiges Interesse und seine hilfreichen Anstöße zu meiner Doktorarbeit. Die Arbeitsbedingungen und die finanzielle Unterstützung waren exzellent!

An dieser Stelle danke ich Herrn Prof. Dr. Rolf D. Schmid sehr herzlich für die Begleitung und die Begutachtung meiner Dissertation. Er war mir ab Beginn meines Studiums ein großer Lehrer; bin ich doch schon seit dem 3. Semester meines Studiums mit seinem Institut für technische Biochemie an der Universität Stuttgart verbunden, zunächst als Hilfswissenschaftler, später bei der Durchführung meiner Diplomarbeit und nun meiner Doktorarbeit.

Bei Frau Prof. Franziska Wollnik bedanke ich mich recht herzlich für das Interesse und die Bereitschaft, diese Arbeit seitens des Instituts für Biologie, Abteilung Tier- und Humanphysiologie zu begutachten.

Meinem Gruppenleiter Prof. Dr. Uli M. Zanger verdanke ich das höchst interessante Thema meiner Doktorarbeit. Ich bin ihm zu Dank verpflichtet für die kreativen Ideen und die anregenden und reflektierenden Gespräche über meine Ergebnisse, die zum wesentlichen Fortgang meiner Doktorarbeit beitrugen. Ebenso danke ich ihm für die Ermöglichung meiner Teilnahme an mehreren internationalen Kongressen. Er war mir immer ein guter Mentor - das weiß ich sehr zu schätzen!

PD. Dr. Till T. Bachmann sende ich meine herzlichen Grüße und meinen Dank an seine neue Wirkungsstätte in Edinburgh/UK. Er war mir stets ein guter Lehrer bei der Entwicklung von Mikroarrays sowohl in meiner Diplom- als auch während meiner Doktorarbeit.

Mein besonderer Dank gilt meinen guten Kolleginnen Dr. Kathrin Klein und Dr. Tanja Saussele für die Zeit, die sie sich immer wieder genommen haben, um meine Ergebnisse mit mir zu diskutieren und für ihre guten wissenschaftlichen Tipps als auch den motivierenden Worten in der „heißen“ Phase.

Für die Tipps und Ratschläge im Labor möchte ich mich bei den technischen Assistenten Igor Liebermann, Britta Klumpp und Andrea Zwicker besonders bedanken – die interessanten Gespräche mit euch, die auch über den Laboralltag hinausgingen werden mir fehlen!

Meinen Büro-Kollegen Dr. Tanja Saussele, Dr. Elke Schäffeler, Dr. Alexander Teml, Susanne Karner und Jessica Rieger teile ich mit, dass wir echt ein klasse Team waren – herzlichen Dank für die bereichernden Gespräche und den Gossip aus dem IKP, den wir zusammen analysiert haben. Ich werde mit euch freundschaftlich verbunden bleiben. Pflegt mir weiterhin „meinen“ Hibiskus!

Ich danke allen (z.T. auch schon ehemaligen) Mitgliedern der Biosensorgruppe am ITB: PD Dr. Till Bachmann, Dr. Timo Barth, Beate Rössle-Lorch, Dirk Leinberger, Dr. Verena Grimm, Susi Münch, Dr. Holger Schulze und Kristina Knösche, sowie den Mitgliedern der kompletten Arbeitsgruppe Pharmakogenomik/Molekular- und Zellbiologie im IKP Prof. Dr. Matthias Schwab, Prof Dr. Uli Zanger, Julia Blievernicht, Diana Feidt, Ana Gomes, Dr. Kathrin Klein, Dr. Reinhold Kerb, Britta Klump, Igor Liebermann, PD Dr. Anne Nies, Dr. Claudia Toscano, Jessica Rieger, Stefan Riedmeier, Dr. Elke Schäffeler, Dr. Shaojun Shi, Dr. Alexander Teml, Dr. Miia Turpeinen und Dr. Jörg Zukunft. Besonders herausheben will ich hier meine Mitdoktorandin Julia Blievernicht. Es war toll mit dir zu dissertieren!

Ich will drei Einrichtungen unvergessen machen, die in den drei Jahren zu meiner Routine gehörten und so manchen Tag gerettet haben: einmal unsere Mittagsrunde mit Tanja, Claudia, Susi, Jessica, Alexander, Ana und Stefan, zweites die große Kaffeerunde, die auch zu der arbeitsgruppenübergreifenden Kommunikation beitrug - hier die Namen aller aufzuzählen würde aber leider den Rahmen sprengen und drittens dem Koffein welcher mich auch nach zehn Stunden im Labor noch gepuscht hat. Danke euch allen!

Schließlich danke ich meinen guten Freunden und davon namentlich meiner Mitstudentin Rosemarie Schäfer und meinem besten Freund Bo Wang sowie meinen Schwestern Beate und Julia für all ihren Rat und ihren Rückhalt. Besonderen Dank gebührt meinen Eltern Gerda und Friedrich Hofmann. Ihnen will ich meine besondere Anerkennung aussprechen für ihre vollste Unterstützung während meiner Schulzeit und meines Studiums. Ihre freundschaftliche, moralische und finanzielle Hilfe bot mir immer größten Rückhalt.

9 Curriculum vitae - Lebenslauf

Name:	Marco Hans Hofmann
Adresse:	Gutbrodstraße 57, 70193 Stuttgart
Nationalität:	deutsch
Geburtsdatum/ Ort:	13. Mai 1978/ Weiden i. d. Oberpfalz

Jan. 2008 **Wissenschaftlicher Mitarbeiter am Institut für Medizinische Virologie, Universität Zürich, Zürich, Schweiz** bei Prof. Dr. K. Mölling und PD Dr. G. Radziwill

Jun. 2004 - Dez. 2007 **Doktorarbeit am Dr. Margarete Fischer-Bosch Institut für klinische Pharmakologie (IKP), Stuttgart in Kooperation mit dem Institut für Technische Biochemie (ITB) der Universität Stuttgart**, Arbeitsgruppe: Pharmakogenomik, betreut von Prof. Dr. M. Schwab (IKP), Prof. Dr. R.D. Schmid (ITB), Prof. Dr. U.M. Zanger (IKP) und PD T.T. Bachmann (ITB) zum Thema: „Microarray and Molecular Genetic Analysis of Aberrant Splicing in Human Drug Metabolizing Cytochromes P450 CYP2D6 and CYP2B6"

Okt. 2001- Mai 2004 **Studium der technischen Biologie an der Universität Stuttgart**
Abschluss als Diplom-Biologe (technisch orientiert)

Sept. 2003 - Mai 2004 **Diplomarbeit am Institut für technische Biochemie der Universität Stuttgart in Kooperation mit dem Dr. Margarete Fischer-Bosch Institut für klinische Pharmakologie**, Arbeitsgruppe: Biosensoren, unter der Leitung von Prof. Dr. R.D. Schmid (ITB), Prof. Dr. M. Schwab (IKP) und PD T.T. Bachmann (ITB) zum Thema: „Detection of TPMT (Thiopurine S-Methyltransferase) SNPs in exon 10 using DNA chip array technology"

Okt. 2001 - Aug. 2002 **Auslandsaufenthalt und Durchführung der Studienarbeit am Nationalen Mikroelektronik Forschungszentrum (NMRC = heute Tyndall National Institute) an der Universität Cork in Irland**, Arbeitsgruppe: bioanalytische Mikrosysteme unter der Leitung von Dr. H. Berney zum Thema: „Development of a miniaturised polymerase chain reaction system", gefördert durch ein Leonardo da Vinci Stipendium

Aug. 2001 - Okt. 2001 **Industriepraktikum in Irland** unter der Leitung von Dr. M.M. Sheehan am Nationalen Mikroelektronik Forschungszentrum (NMRC = heute Tyndall National Institute) in Cork (eine Kooperation mit dem Sister's of Mercy Hospital in Cork) zum Thema: „Characterisation and Investigation of the Sterilisation Effect of Flexible Electrodes"

Sept.1997 - Mai 2000 **Studium der technischen Biologie an der Universität Stuttgart**
Abschluss des Vordiploms

Sept. 1988 - Juli 1997 **Allgemeine Hochschulreife, Kepler-Gymnasium Weiden i. d. Oberpfalz**

Sept. 1984 - Aug. 1988 **Grundschule Weiherhammer**

10 Publication List - Publikationsliste

- U.M. Zanger and M.H. Hofmann, Polymorphic Cytochromes P450 CYP2B6 and CYP2D6: Recent Advances on Single Nucleotide Polymorphisms Affecting Splicing. Acta Chim. Slov. 2008, 55(1), 38-44

- M.H. Hofmann, J.K. Blievernicht, K. Klein, T. Saussele, E. Schaeffeler, M.Schwab, U.M. Zanger, Aberrant Splicing Caused by Single Nucleotide Polymorphism c.516G>T [Q172H], a Marker of CYP2B6*6, Is Responsible for Decreased Expression and Activity of CYP2B6 in Liver. J Pharmacol Exp Ther. 2008, 325(1), 284-92

- U.M. Zanger, K.Klein, T.Saussele, J.Blievernicht, M.H. Hofmann and M. Schwab. Polymorphic *CYP2B6*: molecular mechanisms and emerging clinical significance. 2007, Pharmacogenomics, 2007 Jul; 8(7), 743-59

- M. Hofmann, A. Akkoyun Rowan Flynn, Alan Mathewson, Helen Berney and Michelle M. Sheehan. Development of PCR conditions in a Silicon Microreactor DNA Amplification Device. 2004, Intern. J. Environ. Anal. Chem., 84(11), 821-833

Konferenzvorträge

- M.H. Hofmann, E. Schaeffeler, V.U. Grimm, U.M. Zanger, T.T. Bachmann and M. Schwab. Application of a DNA microarray for genotyping of functional SNP in the Thiopurine S-methyltransferase (TPMT) gene. Pharmacogenomics, Hinxton, Cambridge, Great Britain, 14-18.09.2005

- M. Hofmann, A. Akkoyun, A. Mathewson and H. Berney. Detection of microorganisims in a microfabricated silicon thermal cycler. 6th Euroconference on Environmental Analytical Chemistry, Brüssel, Belgium 16-22.10.2002

Posterveröffentlichungen

- M.H. Hofmann, E. Schaeffeler, K. Klein, M. Eichelbaum, M. Schwab, T.T. Bachmann and U.M. Zanger. A splicing microarray strategy for profiling alternative patterns from functional genes and corresponding pseudogenes. Statuseminar of Chiptechnology, Frankfurt, Germany, 01-02.02.2007

- M.H. Hofmann, C. Toscano, M. Eichelbaum, M. Schwab, T.T. Bachmann and U.M. Zanger. Developement of an oligonucleotide-based microarray to detect CYP2D splice variations. Microsomes and Drug Oxidations, Budapest, Hungary, 03-07.09.2006

- C. Toscano, M.H. Hofmann, S. Raimundo, K. Klein, E. Schaeffeler, M. Eichelbaum, M. Schwab, U.M. Zanger. Analysis of splice variations associated with the *CYP2D6*41* intermediate metabolizer phenotype. Alternative Transcript Diversity – Data, Biology and Therapeutics, Hinxton, Cambridge, Great Britain, 22-23.09.2004

- H. Berney, P. O'Brien, J. West, B. Lillis, M. Sheehan, A. Akkoyun, M. Hofmann, H. Ke, E. Hurley, W. A. Lane and A. Mathewson. The use of Microtechnology in Biotechnology and Analytical chemistry. Analytical conference IT Tallaght., Ireland 04-05.04.2002

- A. Akkoyun, M. Hofmann, A. Mathewson and H. Berney. Towards Detection of Microorganisms in Micro-fabricated Silicon thermal Cycler. Analytical conference IT Tallaght, Ireland, 04-05.04.2002

◆ M. Hofmann, A. Akkoyun, A. Mathewson and H. Berney. Optimisation of Polymerase chain reaction in micro-fabricated silicon thermal cycler. Analytical conference IT Tallaght, Ireland, 04-05.04.2002

◆ J. Piggott, H. Berney, E. Hurley, J. Clair, M. Hofmann, F. Stam, A. Morrissey and M. Sheehan. Flexible Plannar Electrodes for use in Wound Sterilization. IEEE Engineering in Medicine and Biology Society and the Biomedicine Engineering Society, Texas, USA 23-26.10.2002