

DNA Microarray Based Gene Expression Profiling in Human Hepatocyte Cells to Serve as a Basis for Dynamic Modelling of the Human Liver - a Systems Biology Approach

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

Thomas Reichart

aus Sindelfingen

Hauptberichter: Prof. Dr. Rolf D. Schmid, MBA

Mitberichter: Prof. Dr. Dr. h.c. Matthias Reuss

Tag der mündlichen Prüfung: 10. Juni 2008

Institut für Technische Biochemie der Universität Stuttgart

2008

TABLE OF CONTENTS

TABLE OF CONTENTS	2
EIDESSTATTLICHE ERKLÄRUNG	5
ACKNOWLEDGMENT	6
ZUSAMMENFASSUNG.....	7
ABSTRACT	15
ABBREVIATIONS	21
1 INTRODUCTION	24
1.1 SYSTEMS BIOLOGY - A NEW APPROACH	24
1.2 "HEPATOSYS" - A COMPUTATIONAL APPROACH TO UNDERSTAND FUNCTIONAL ASPECTS OF THE LIVER	31
1.3 XENOBIOTIC METABOLISM	34
1.4 MICROARRAY TECHNOLOGY.....	40
1.4.1 <i>Microarrays in Drug Discovery, Drug Development and Future Applications</i>	50
1.4.2 <i>Basics of Microarray Technology</i>	55
1.4.3 <i>Data Quantification and Interpretation of Microarray Experiments</i>	66
1.4.4 <i>Identifying Differentially Expressed Genes</i>	69
1.5 OBJECTIVE OF THIS THESIS	76
2 MATERIALS AND METHODS	78
2.1 CHEMICALS, ENZYMES AND KITS	78
2.2 HARDWARE EQUIPMENT	79
2.3 SOFTWARE TOOLS AND INTERNET RESOURCES	80
2.4 BUFFERS AND SOLUTIONS.....	81
2.4.1 <i>Complex media</i>	81
2.4.2 <i>Antibiotics</i>	82
2.5 MEDIA FOR CULTIVATION OF PRIMARY HUMAN HEPATOCYTES	82
2.6 PRIMARY HUMAN HEPATOCYTE CELLS, BACTERIAL STRAINS AND PLASMIDS USED IN THIS WORK	83
2.6.1 <i>Cultivation of Bacterial Cells</i>	86
2.6.2 <i>Conservation of Strains</i>	87
2.7 CELL CULTIVATION AND INDUCTION OF PRIMARY HEPATOCYTES.....	87

2.7.1	<i>Bringing Primary Hepatocytes into Culture after Transport</i>	87
2.7.2	<i>Induction Experiments of Primary Hepatocytes</i>	88
2.7.3	<i>Harvesting of Primary Hepatocytes</i>	89
2.8	METHODS OF MOLECULAR BIOLOGY	90
2.8.1	<i>Working with RNA</i>	90
2.8.2	<i>Purification of Nucleic Acids</i>	91
2.8.3	<i>Quality Control of Nucleic Acids</i>	92
2.8.4	<i>Precipitation of Nucleic Acids</i>	94
2.8.5	<i>RNA Amplification</i>	94
2.8.6	<i>Digestion of DNA</i>	100
2.8.7	<i>Separation of Nucleic Acids Using Agarose Gel Electrophoresis</i>	101
2.8.8	<i>Preparation and Transformation of Competent E. coli</i>	104
2.8.9	<i>Reverse Transcription</i>	106
2.8.10	<i>Quantitative Real-time PCR</i>	108
2.9	PROTEIN BIOCHEMICAL METHODS	114
2.9.1	<i>Protein Precipitation with Acetone</i>	114
2.9.2	<i>SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)</i>	114
2.9.3	<i>Semi-dry Protein Transfer for Western Blotting</i>	119
2.9.4	<i>Immunological Detection of Proteins on Nitrocellulose Filters</i>	120
2.10	METHODS USING DNA MICROARRAYS	122
2.10.1	<i>Labeling of aRNA/cDNA</i>	122
2.10.2	<i>Indirect Labeling of Amplified (a)RNA</i>	122
2.10.3	<i>Direct Labeling During Reverse Transcription</i>	123
2.10.4	<i>Indirect Labeling During Reverse Transcription</i>	125
2.10.5	<i>Hybridization and Washing</i>	127
2.10.6	<i>Detection of Indirectly Labeled Targets</i>	129
2.10.7	<i>Scanning of Arrays</i>	129
2.10.8	<i>Extraction of Raw Data and Picture Analysis (Quantification)</i>	130
2.10.9	<i>Evaluation of DNA Microarrays and Data Analysis</i>	131
2.10.10	<i>Analysing Sub-Genome Data using the Eppendorf DualChip® Software</i>	132
2.10.11	<i>Analysing Full-Genome Data Using "Refiner" and "Analyst" from Genedata</i>	138
2.10.12	<i>Literature Mining Using Bibliosphere from Genomatix</i>	147
3	RESULTS	152
3.1	COMPARISON OF MICROARRAYS FROM DIFFERENT SUPPLIERS	153
3.2	EXPERIMENTAL SETUP OF TIME-SERIES EXPERIMENTS	162
3.3	RNA ISOLATION AND QUALITY CONTROL	165
3.4	VERIFICATION OF INDUCTION OF CELL CULTURES	173
3.4.1	<i>Quantitative Real-Time PCR (qRT-PCR)</i>	173

4 Introduction

3.4.2	<i>Western Blotting</i>	177
3.5	REVERSE TRANSCRIPTION (RT), AMPLIFICATION AND LABELING OF MRNA.....	178
3.5.1	<i>Direct Labeling with Fluorescent Dyes</i>	178
3.5.2	<i>Amplification and Indirect Labeling Using Biotin Modified Nucleotides</i>	181
3.6	HYBRIDIZATION OF ARRAYS AND QUALITY CONTROL.....	185
3.6.1	<i>Quality Control for Sub-Genome Arrays from Eppendorf</i>	185
3.6.2	<i>Quality Control for Full-Genome Arrays from Affymetrix</i>	189
3.7	INVESTIGATION OF TECHNICAL- AND BIOLOGICAL REPLICATES.....	191
3.8	DETERMINATION OF DIFFERENTIALLY EXPRESSED GENES.....	194
3.8.1	<i>Whole Genome Approach</i>	195
3.8.2	<i>Gene Expression Profiling Based on Cellular Function</i>	234
3.8.3	<i>Reconstruction of Networks Using a Combined Literature Mining and Microarray Analysing Approach (LMMA)</i>	247
3.8.4	<i>Interaction of Xenobiotic with Endogenous Metabolism Based on Nuclear Receptor Interaction</i>	263
3.8.5	<i>Time-Series Approach with Sub-Genome Arrays</i>	267
3.8.6	<i>Time-Series Approach Based on Real-Time PCR</i>	287
4	DISCUSSION	289
4.1	METHODICAL CONSIDERATIONS.....	289
4.1.1	<i>Obtaining Human Samples for Microarray Gene Expression Profiling</i>	290
4.1.2	<i>RNA Isolation, Amplification and Labeling</i>	291
4.1.3	<i>Comparison of Expression Data from Different Microarray Platforms</i>	293
4.1.4	<i>Network Reconstruction Based on Microarray- and Literature Data</i>	297
4.2	PHYSIOLOGICAL CONSIDERATIONS.....	297
4.2.1	<i>Control of Cytochrome P450 Enzyme Family (CYPs) and 5-Aminolevulinate Synthase (ALAS1) Gene Transcription by Nuclear Receptors</i>	301
4.2.2	<i>Regulation of Bile Acid Metabolism by Nuclear Receptors</i>	311
4.2.3	<i>Glucose- and Fat Metabolism - the Role of Nutritional Regulation</i>	316
4.2.4	<i>Induction of Metallothionein Proteins</i>	320
4.2.5	<i>Down-regulation of Cadherin Cell Adhesion Receptor</i>	323
4.2.6	<i>Effects of Inflammatory Cytokines on Expression of Cytochrome P450 Enzymes</i>	324
4.3	CONCLUSION.....	326
5	LITERATURE	329
6	APPENDIX	347
	CURRICULUM VITAE	365

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Literatur angefertigt habe.

Stuttgart, den

Acknowledgment

Bedanken möchte ich mich an erster Stelle recht herzlich bei Herrn **Prof. Dr. Rolf D. Schmid, MBA** für die Bereitstellung des herausfordernden, hochinteressanten Themas, seinen hilfreichen Anregungen und seiner ständigen Diskussionsbereitschaft, die neben den hervorragenden Bedingungen an seinem Institut viel zum Gelingen der vorliegenden Arbeit beigetragen haben.

Herrn **Prof. Dr.-Ing. Matthias Reuss** danke ich für sein Interesse an der vorliegenden Arbeit und die Bereitschaft, diese als Zweitgutachter zu beurteilen.

Des Weiteren möchte ich **Dr. Stefan Lange** und insbesondere **Dr. Steffen Maurer** für die Durchsicht der Arbeit und die Betreuung danken.

Bedanken möchte ich mich auch bei meinem Freund und Kollegen **Jan Pfeffer**, mit dem ich gemeinsam alle Höhen und Tiefen einer Doktorarbeit durchlebt habe.

Dr. Jürgen Dippon und Herrn **Stefan Winter** danke ich für die Mithilfe bei der Planung und Durchführung von zwei Microarray-Workshops in Stuttgart, die Zusammenarbeit im Rahmen der Auswertung der Array-Daten und die statistische Begutachtung der Ergebnisse.

Dr. Ulrich Zanger und **Dr. Kathrin Klein** danke ich für die Zusammenarbeit und Hilfestellung bei medizinischen Fragen und die Unterstützung bei der Durchführung von real-time PCR basierten Messungen.

Prof. Dr. Andreas Nüssler danke ich sehr herzlich für die Bereitstellung von Probenmaterial (Zellkulturen) für meine Untersuchungen.

Gabriele Vacun möchte ich für die die Einweisung in die Kultivierung von primären Zellkulturen und die gemeinsame Durchführung danken.

Herrn Prem Kumar Murugan danke ich für die gute Zusammenarbeit und die interessanten Einblicke in die Modellierung meiner Expressions-Daten.

Den Kollegen aus "Hepatosys", insbesondere aus dem Stuttgarter Verbund, möchte ich für die vielen interessanten Diskussionen und Anregungen danken.

Den Mitarbeitern von Eppendorf Biochipsystems danke ich für die fachkundige Beratung und Hilfestellung.

Meinen Eltern und meinem Bruder danke ich für die stets vorhandene Unterstützung in allen Bereichen.

Dem Bundesministerium für Bildung und Forschung (BMBF) danke ich für die Finanzierung der Arbeit.

Zusammenfassung

Der Mensch ist im Laufe seines Lebens einer Vielzahl von Fremdstoffen ausgesetzt. Diese reichen von Nahrungsmitteln über pharmazeutische Substanzen bis hin zu Giftstoffen aus der Umwelt. Die Evolution hat mit einem ausgeklügelten Enzymsystem dafür gesorgt, dass der menschliche Organismus im Stande ist, mit all diesen Substanzen, die häufig völlig neuartig sind, korrekt umzugehen. Diese Tatsache bedarf nicht nur einer ungemeinen Fülle an Enzymen mit einer breiten Substratspezifität, sondern auch eines ausgeklügelten Regulationsnetzwerkes, das wesentlicher Bestandteil der Untersuchungen in der hier vorliegenden Arbeit ist.

Zahlreiche der Detoxifizierung von Fremdstoffen zugrunde liegenden Mechanismen sind bis heute ungeklärt. Die Entwicklung neuer analytischer Verfahren, insbesondere im Hochdurchsatzbereich, wie etwa die Microarray Technologie, die die vollständige Beschreibung des Expressionsverhaltens eines Organismus' erlaubt oder die Pyrosequenzierung ganzer Genome in nur wenigen Tagen, eröffnen völlig neue Möglichkeiten der Datenerhebung und Analyse. Hinzu kommt die Vernetzung von Wissen aus verschiedenen Bereichen durch interdisziplinäre Forschungs Kooperationen, die ein Zusammenwirken verschiedener Disziplinen wie der Informatik, den Systemwissenschaften, der Regelungstechnik, der Chemie und der Biologie erreichen. Diese neue fächerübergreifende, ganzheitliche Betrachtungsweise fasst man unter dem Begriff Systembiologie zusammen. Ursprünglich an technischen Systemen gewonnene Erkenntnisse können in solchen Ansätzen zu einem besseren Verständnis biologischer Stoffwechsel- und Regulationsprozesse führen. Komplexe technische Prozesse sind mit denen der Biologie häufig eng verwandt. Modularität, hierarchische Strukturierung der Regulation, Redundanz und Diversität gehören offensichtlich zu den unverzichtbaren Gestaltungsmerkmalen der Systeme und Prozesse in beiden Bereichen.

Die Grundlage für ein tieferes Verständnis des komplexen Geschehens der Detoxifizierung von Fremdstoffen sind experimentelle Daten. Ziel der vorliegenden Arbeit war die Erhebung dieser Daten im Rahmen der Messung von Botenribonukleinsäure (mRNS) zu verschiedenen Zeitpunkten nach Induktion mit

einem Fremdstoff. Die Menge an vorhandener mRNA erlaubt Aussagen über die Veränderung der Genexpression. Der Großteil der Umwandlung und Entgiftung von Stoffen in Säugetieren erfolgt in der Leber. Um Aussagen über die zu Grunde liegenden Mechanismen beim Menschen treffen zu können, wurden die Untersuchungen an primären menschlichen Leberzellkulturen (Hepatozyten) durchgeführt. Als Modellsubstanz für einen Fremdstoff wurde das macrocyclische Antibiotikum Rifampicin gewählt.

Teilaspekte der Induktion primärer Leberzellen mittels Rifampicin wurden in der Vergangenheit bereits eingehend untersucht: so ist bekannt, dass Rifampicin ein starker Induktor für Cytochrom P450 Monooxygenasen (CYPs) ist (Zilly *et al.*, 1977; Fromm *et al.*, 1996; Niemi *et al.*, 2003). Rifampicin bindet mit hoher Affinität an den menschlichen Glucocorticoidrezeptor, der eine Vielzahl von Genen reguliert, unter anderem auch Interleukin und CYP codierende Gene (Calleja *et al.*, 1998). Da Rifampicin gleichzeitig durch CYP3A Enzyme verstoffwechselt wird, ermöglicht die Induktion mit Rifampicin Einblicke in Rückkopplungsmechanismen einer Autoinduktion bzw. eines negativen Feed-back. Aus diesen Gründen wurde Rifampicin als Modellsubstrat zur Modellierung eines Regulationsnetzwerks einer Leberzelle basierend auf Expressionsdaten gewählt.

Zunächst wurde das bereits vorhandene Wissen analysiert und musste zum Teil in seiner Übertragbarkeit auf den Menschen verifiziert werden, da zahlreiche der vorherigen Arbeiten auf Experimenten mit nicht menschlichen Hepatozyten basierten. Da die Übertragbarkeit von Daten, die mit alternativen Modellsystemen gewonnen wurden, nicht geklärt ist, ist man zur Analyse der menschlichen Leber auf Experimente mit humanen Zellkulturen mit all ihren Schwierigkeiten, wie dem geringen Probenaufkommen und der hohen individuellen Varianz der Spender, angewiesen. Ein weiterer Nachteil der bisher vorliegenden Untersuchungen besteht darin, dass sich die Mehrheit der durchgeführten Analysen auf bestimmte Teilbereiche der Zellantwort beschränken. So wurde beispielsweise der Einfluss von Fremdstoffen und natürlichen Steroiden auf CAR und PXR aus der Maus und dem Menschen untersucht (Moore *et al.*, 2000), die Interaktion von PXR und Co-Aktivatoren (Li and Chiang, 2006) oder die Pharmaka induzierte Herunterregulierung der Gluconeogenese durch Aktivierung von PXR (Kodama *et al.*, 2007), die auch in dieser Arbeit nachgewiesen werden konnte.

Die hier vorgestellte Arbeit bietet zum ersten Mal einen ganzheitlichen Ansatz zur Untersuchung der zeitaufgelösten Genexpression in humanen Leberzellkulturen nach einem Stimulus mit Rifampicin. Die Daten aus der das gesamte menschliche Genom umfassenden Analyse mittels Desoxyribonukleinsäure-Microarrays (DNS-Microarrays) wurden durch zusätzliche auf Echtzeit-Polymerasekettenreaktion (qRT-PCR) basierende Experimente mit einer Auswahl an Genen ergänzt und verfeinert.

Ein weiteres Ziel der hier vorliegenden Arbeit bestand im Vergleich geeigneter analytischer Verfahren zur Messung der transkriptionellen Veränderung in primären Leberzellkulturen und deren Etablierung. Getestet wurden drei Plattformen der Firmen Affymetrix, Agilent und Eppendorf in Bezug auf Leistungsumfang, Handhabung, Preis, Reproduzierbarkeit und Qualität der Ergebnisse. Die etablierten Methoden wurden, wie Eingangs skizziert, zur spezifischen Analyse der Regulation von Cytochrom P450 3A4 (CYP3A4) auf Transkriptebene sowie zur globalen Untersuchung der Veränderung der Gen Expression des Transkriptoms herangezogen.

Humane Leberzellkulturen wurden von Patienten gewonnen, die sich einer Leberoperation unterzogen hatten. Die Induktionsexperimente erfolgten im Rahmen einer Kooperation mit dem Institut für Bioverfahrenstechnik (IBVT) der Universität Stuttgart. Die aus dem Expressionsprofiling gewonnenen Daten wurden im Rahmen dieser Kooperation an das IBVT weitergegeben und dienten dort als Grundlage für die Rekonstruktion des Regulationsnetzwerkes von CYP3A4.

Die Induktion der primären Leberzellkulturen erfolgte durch das bereits als Induktor für eine Vielzahl von Genen charakterisierte Antibiotikum Rifampicin. Um die Dynamik der Regulation besser abbilden zu können, wurden jeweils mindestens drei Zeitpunkte nach Induktion analysiert. Zu jedem Zeitpunkt wurden neben den induzierten Proben auch solche Proben genommen und analysiert, die als Kontrolle nur DMSO enthielten, da Rifampicin in DMSO gelöst verwendet wurde.

Die auf zwei systembiologischen Prinzipien des Erkenntnisgewinns basierenden Analyseverfahren lassen sich wie folgt darstellen: im sogenannten "Top-Down" Ansatz werden aus der allumfassenden Beschreibung des Gesamtsystems deduktiv neue Erkenntnis gewonnen. Dieser Ansatz wurde mit der Analyse der Genexpression unter Verwendung von DNS Microarrays der Firma Affymetrix verfolgt, die die parallele Analyse von > 20.000 menschlichen Genen erlaubt. In der vorliegenden Arbeit wurde

die Expression des gesamten Genoms an drei Patientenproben zu jeweils drei Zeitpunkten (6, 24 und 72 h nach Induktion) in induzierten- und Kontrollproben untersucht. Aus den erhaltenen Daten wurden im Rahmen dieser Arbeit und in Kooperation mit dem IBVT Erkenntnisse über neue, an der Regulation der Induktion von CYP3A4 durch Rifampicin beteiligte Interaktionen (PXR mit HNF4 α sowie den Koaktivatoren SRC-1 und PGC-1) gewonnen.

Parallel wurde, basierend auf bereits vorhandenem biologischem Wissen, ein induktiver "Bottom-Up" Ansatz verfolgt: dazu wurde durch eine Vorauswahl von Genen die Fragestellung untersucht, welchen Einfluss die Induktion von CYP3A4 durch Rifampicin auf den Stoffwechsel und die Regulationsnetzwerke der Leberzelle hat. Diese kleinere Einheit an Genen konnte dann detaillierter analysiert werden als es für die Gesamtheit aller Gene möglich wäre. Experimentell wurde dieser Ansatz mit subgenomischen DNS-Chips und quantitativer Echtzeit-Polymerasekettenreaktion (qRT-PCR) durchgeführt. Insbesondere der geringere Kostenaufwand erlaubte eine zeitlich besser aufgelöste Analyse der Genexpression mit einer höheren Anzahl an Proben: das dynamische Expressionsmuster wurde 6, 12, 16, 20, 24, 28 und 72 Stunden nach Induktion mittels DNS-Chips und 0,5, 1, 4, 8, 12, 16, 24, 30, 36, 42 und 48 Stunden nach Induktion basierend auf Echtzeit-PCR Messungen bestimmt.

Während der "Top-Down" Ansatz die Identifikation interessanter Gene ermöglicht, verfolgt der "Bottom-Up" Ansatz die Verfeinerung des Wissens über zuvor bereits identifizierte Bereiche der zellulären Antwort.

Zur Analyse der Genexpression in Leberzellen wurden die Zellen zu bestimmten Zeitpunkten nach Induktion geerntet und die Gesamt-RNS aus den Induktions- und Kontroll-Proben isoliert. Einer der wichtigsten Faktoren für den Erfolg von Experimenten, die das Gen Expressionsmuster einer Zelle zum Zeitpunkt der RNS Extraktion widerspiegeln sollen, ist die Integrität der RNS Moleküle. Daher wurde ein besonderes Augenmerk auf die Überprüfung der RNS Qualität gelegt. Eine auf Mikrofluidik basierende, in Kapillaren durchgeführte elektrophoretische Trennung der RNS mit anschließender laserinduzierter Fluoreszenz Detektion war die Grundlage der etablierten Methode zur Bestimmung der "RNA integrity number" (RIN). Die RIN erlaubte die vergleich- und reproduzierbare Bestimmung der Integrität von RNS und damit deren Verwendbarkeit für nachfolgende Array- oder qRT-PCR Experimente. Die

mit dem Agilent 2100 Bioanalyzer durchgeführten Analysen ermöglichten die quantitative Bestimmung der Qualität einer RNS Probe und somit auch den objektiven Ausschluss von Proben minderer Qualität. Die Ausschlussgrenze wurde für RIN-Werte $< 7,5$ definiert. RNS Proben mit einer $RIN > 7,5$ wurden je nach verwendetem Verfahren entweder in cDNS umgeschrieben oder amplifiziert. cDNS-Synthese war die Grundlage für quantitative Bestimmungen mittels Echtzeit-PCR oder Chipexperimenten mit direkter Markierung. Alternativ dazu wurde die mRNA in den meisten Fällen jedoch zunächst amplifiziert und dann für Chipexperimente mit indirekter Markierung verwendet (Affymetrix und Eppendorf). Da der Unterschied zwischen der Genexpression in einer induzierten Probe im Vergleich zu einer Kontrollprobe zu verschiedenen Zeitpunkten untersucht werden sollte, wurde zur Hybridisierung ein "time matched control design" gewählt. Dabei wurde parallel zu jeder induzierten Probe eine nicht induzierte Vergleichsprobe hybridisiert. Das Verhältnis der beiden Werte ermöglicht die Bestimmung der differentiellen Genexpression zu den entsprechenden Zeitpunkten.

Die wesentlichen Ergebnisse der vorliegenden Arbeit sind wie folgt: 24 h nach Induktion primärer humaner Leberzellen mit Rifampicin konnte eine regulierte Transkription infolge des Stimulus nachgewiesen werden. Gene des Xenobiotika-metabolismus wie ATP-abhängiger Transporter der Unterfamilie B (MDR/TAP) Nummer 1 (ABCB1), Cytochrome P450 Enzyme (CYPs), der lösliche Transporter der Familie 1 Nummer 4 (SLC1A4) und Familie 7 Nummer 1 (SLC7A1) wurden vermehrt transkribiert. Signifikant hochreguliert wurde auch die Expression der Aminolävulinat delta-Synthase 1 (ALAS1). ALAS1 ist das regulierte Eingangs- und Schlüsselenzym der Häm Biosynthese. Die Koregulierung der Expression dieses Enzyms mit den P450-Enzymen ist dadurch erklärbar, dass das Häm bei der Neusynthese der P450-Monooxygenasen für deren aktives Zentrum benötigt wird.

Zusätzlich wurde eine Gruppe von Metallothionein Proteinen als signifikant hochreguliert identifiziert, die möglicherweise bei der Entgiftung von im Stoffwechsel anfallenden radikalen Sauerstoffspezies beteiligt sind. Diese Spezies können als Nebenprodukte der Hydroxylierung durch P450-Monooxygenasen entstehen. Des Weiteren wurde eine Herunterregulierung der an der Glykogenspeicherung beteiligten

Gene Glykogen Synthase 2 (Leber, GYS2) und Phosphoenolpyruvat Carboxykinase 1 (PCK1) gefunden. Dies war ein sehr interessanter Befund, da über die Interaktion des Kohlenstoffmetabolismus' mit dem Fremdstoffmetabolismus diskutiert wird. Der Ernährungszustand hätte demzufolge, neben der genetischen Ausstattung, dem Alter und dem Geschlecht des Patienten einen erheblichen Einfluss auf den (Xenobiotika-) Metabolismus. In diesem Zusammenhang wurde der Einfluss der Induktion auf den Lipidstoffwechsel, den Gallensäurestoffwechsel und den Bilirubinauf- und -abbau näher untersucht und charakterisiert. Diese Stoffwechselfunktionen finden auch in der Leber statt und werden zum Teil von den gleichen metabolischen Enzymen (CYP3A4) katalysiert, bzw. den gleichen Rezeptoren (PXR) reguliert.

Interessanterweise wurden in der Mehrheit der Experimente die Hauptkomponenten der Regulation wie der Hepatozyten Kernfaktor Alpha (HNF4 α), beziehungsweise die Kernrezeptoren PXR und CAR, als leicht herunterreguliert identifiziert. Die starke Induktion der Zielgene dieser Rezeptoren, die parallel gezeigt werden konnte, deutet darauf hin, dass keine Erhöhung der Expression erforderlich ist, um auf einen Stimulus (Rifampicin) zu reagieren. Möglicherweise reicht die vorhandene Menge an Rezeptorprotein aus, um durch Bindung an die Promotorregion der Zielgene deren Transkription zu veranlassen. Um diese Hypothese zu überprüfen, werden in einer Anschlussarbeit bereits weitergehende Untersuchungen mittels Biacore Messungen und Analysen durch Chromatin Immunopräzipitation auf einem Chip (ChIP-Chip) durchgeführt. Diese sollen Aufschluss über Bindungen der Kernrezeptoren mit ihren Liganden als auch mit ihrer Ziel-DNA und die daraus resultierende Induktion geben.

Die Analyse mittels sub-genomischer DNA Arrays zeigte erhebliche interindividuelle Unterschiede. Insbesondere die Induktion von Cytochrom P450 Enzymen (CYPs), der Glutathion S-Transferase A2 (GSA, GSTA-2), dem ATP-Transporter (ABCC1) und Cyclin B1 (CCNB1) sowie Enoyl-Co-Enzym A, Hydratase/3-hydroxyacyl Co-Enzyme A Dehydrogenase (EHHADH) zeigte deutliche Schwankungen der Genexpression in Proben verschiedener Patienten. Dennoch konnten zahlreiche vergleichbar induzierte als auch reprimierte Gene über verschiedene Patientenproben hinweg identifiziert werden. So konnte beispielsweise auf Expressionsebene ein Zusammenhang zwischen

inflammationsgekoppelter CYP3A4-Repression durch das Interleukin 6 (IL-6) gezeigt werden.

Die Induktion von CYP3A4 durch Rifampicin wurde zudem mittels Echtzeit-PCR und auf Proteinebene mittels Western-Blot verifiziert. Der qRT-PCR Assay wurde des Weiteren im Rahmen eines Zeitreihenexperiments angewandt, um einen zeitaufgelösten Verlauf der Genexpression ausgewählter Gene (CYP3A4 und PXR) zu erhalten. Solche zeitlich eng aufgelösten Messungen sind essentielle Grundlage für die dynamische Modellierung. Sie zeigten beispielsweise, dass während des starken Anstiegs der mRNA Menge von CYP3A4 in allen induzierten Zellkulturen die mRNA Menge von PXR in den gleichen Zellen nur eine vergleichsweise geringe und zudem variable Zu- bzw. sogar Abnahme aufwies. Außerdem konnten differenzierte Induktionsverläufe für verschiedene Cytochrom P450 Enzyme beschrieben werden. Diese Vorgehensweise ist jederzeit auf weitere Gene erweiterbar, insbesondere vor dem Hintergrund, dass mittlerweile beinahe das gesamte menschliche Genom als Echtzeit-PCR Assay kommerziell verfügbar ist.

Im Rahmen der in dieser Arbeit durchgeführten Genexpressionsstudien konnten zahlreiche Ergebnisse wissenschaftlicher Arbeiten ergänzt und bestätigt werden. Der Vergleich dreier kommerziell erhältlicher Microarray-Plattformen zeigte, dass die Daten der untersuchten Plattformen vergleichbar reproduzierbare Ergebnisse liefern. Es wurde eine Korrelation von 0,62 zwischen den Daten der Affymetrix- und Agilent Chips sowie eine Korrelation von 0,63 zwischen den Resultaten, die mit Affymetrix und mit Eppendorf Chips erhoben wurden, gefunden. Zudem wurde gezeigt, dass die Interplattform Variabilität für Affymetrix Arrays bei 0,93 und bei Eppendorf Chips bei 0,8 lag. Des Weiteren wurde eine Vielzahl an signifikant differenziell exprimierten Genen, reproduzierbar sowohl mit der Affymetrix als auch der Eppendorf Plattform, nach Induktion durch Rifampicin identifiziert. Diese haben unter anderem Aufgaben im Energiestoffwechsel (Gluconeogenese), in der Häm-Synthese, im Bilirubin-stoffwechsel, im Gallensäurestoffwechsel, im Lipidstoffwechsel und nicht zuletzt im Fremdstoffwechsel und Transport. Die Beobachtung der Antwort der Zellen auf RNA Ebene über die Zeit erlaubte zudem in Kooperation mit dem Institut für Bioverfahrenstechnik neue Einblicke in die Regulation dieses komplexen Geschehens.

Neben der Netzwerkrekonstruktion, basierend auf den das gesamte Genom umfassenden Array Daten, ermöglichen die mit einer höheren zeitlichen Auflösung vorliegenden Daten eine dynamische Modellierung der Expression und damit eine mathematische Abbildung der Antwort der Zelle auf den Stimulus. Basierend darauf sollen in Zukunft auch prädiktive Aussagen über das Verhalten des "System Hepatozyt" möglich sein. DNA Chip basierte Pharmakogenomik und Toxikogenomik wurden kürzlich sowohl durch die "US Food and Drug Administration" (FDA) als auch die "US Environmental Protection Agency" (EPA) als Schlüssel zur Entwicklung personalisierter Medizin identifiziert. Die auf "omics" Daten basierenden, modellgestützten Vorhersagen werden neben der personalisierten Medizin (Diagnose und Therapie) insbesondere auch eine Verbesserung der Medikamentenentwicklung und Risikoabschätzung mit sich bringen.

Abstract

In his lifetime the human body is exposed to a large number of xenobiotics. These include food components, pharmaceutical substances and environmental toxins. Evolution has developed complex enzymatic mechanisms to detoxify these compounds; even if they show little relationship to previously encountered compounds or metabolites which occurs quite often. Detoxification requires not only an immense number of enzymes with broad substrate specificity but also an elaborate regulation. A lot of basic mechanisms of detoxification are still poorly understood. Shedding more light on this regulation is a major part of the work presented here.

The development of new analytical methods especially high throughput technologies e.g. DNA microarrays (recently reviewed in Jayapal and Melendez, 2006; Lettieri, 2006; Wang and Cheng, 2006; Mandruzzato, 2007), or pyrosequencing (Ronaghi, 2001) of whole genomes in a few days open a wide field of new possibilities. Furthermore the attempt to combine innovative, interdisciplinary research collaborations looks quite promising. These collaborations will allow networking of experts from different subjects like computer science, systems-engineering, chemistry and biology. One of the ideas behind systems biology is that the knowledge gained in systems description of technical applications can help to understand biological metabolite- and regulatory networks. Complex technical processes are often closely related to biological ones. Modularity, hierarchical structures of regulation, redundancy and diversity are obviously indispensable constitutive criteria of the systems and processes in both areas.

The basis for a better understanding of the complex detoxification processes of xenobiotics is of course experimental data. One aim of the work presented here was the measurement of mRNA concentrations at different points in time after inducing primary human hepatocyte cultures with a xenobiotic. Based on these measurements changes of gene expression upon induction could be revealed. The main part of transformations and detoxification of xenobiotics in mammals occurs in the liver. To be able to make assumptions about the underlying mechanisms in the human body, experiments were

performed using primary human liver (hepatocyte) cell cultures. As model substance for a xenobiotic the macrocyclic antibiotic Rifampicin was chosen.

Induction of primary human hepatocytes using Rifampicin was already partially studied. It is known that Rifampicin is a strong inducer for cytochrome P450 monooxygenases (CYPs) (Zilly *et al.*, 1977; Fromm *et al.*, 1996; Niemi *et al.*, 2003). Rifampicin binds to the human glucocorticoid receptor with high affinity, which regulates the expression of many genes, including those encoding interleukins that regulate immune response and CYPs (Calleja *et al.*, 1998). As Rifampicin is degraded by CYP3A4, induction with this antibiotic additionally offers insights into positive as well as negative feedback regulation and auto induction mechanisms. Consequently this antibiotic is a perfect model substance for experiments with the goal to deliver data for modelling regulatory networks of the liver cell.

First of all existing publications had to be analysed and verified to be also valid for humans, as most of the previous work was based on experiments using non human hepatocyte cells (Li *et al.*, 2007). There are interesting approaches of using chimera mice with humanized livers. This could allow easier and more reliable comparison of experiments as well as improve availability of samples (Yoshitsugu *et al.*, 2006). However as long as there is no final evaluation on the benefit of this kind of experiments, one has to rely on experiments using human hepatocyte cell cultures for analysing the human liver despite all associated difficulties. These include shortage of samples and high individual variance of samples from different donors.

In addition most of the recent analyses were focused on specific parts and details of the cellular response, e.g. the effects of xenobiotics and natural steroids on mouse and human CAR and PXR (Moore *et al.*, 2000), the interaction of PXR with HNF4alpha, SRC-1, and PGC-1alpha (Li and Chiang, 2006) or the down-regulation of gluconeogenesis by drug activation of PXR (Kodama *et al.*, 2007), also observed in this study.

The work presented here gives for the first time a holistic approach for the measurement of time-dependent messenger ribonucleic acid (mRNA) concentration in liver cell cultures after stimulating with Rifampicin. The obtained deoxyribonucleic acid-microarray (DNA-microarray) data from the whole genome study of the

expression profiles after induction was extended and refined by additional real-time polymerase chain reaction (PCR) based measurements for a selected number of genes. An additional goal of this work was the comparison and establishment of suitable analytical methods to detect transcriptional changes in primary liver cell cultures. Three microarray platforms from Agilent, Affymetrix and Eppendorf were compared with respect to performance, scope, cost, reproducibility and quality of results. The established methods were used in a subsequent step to analyse the regulation of the induction of cytochrome P450 3A4 (CYP3A4) on transcript level and for a holistic exploration of the gene expression of the transcriptome.

Human liver cell cultures were obtained from patients who had to undergo liver surgery. Induction experiments were done in cooperation with the Institute of Biochemical Engineering (IBVT) at the University of Stuttgart. Data gained from the microarray based expression profiling was transferred to the IBVT. This data was the scaffold for the modelling group who reconstructed the regulatory network of CYP3A4 based on the measurements described here.

To get a detailed picture of the dynamics of regulation at least three points in time (6, 24 and 72 h) after induction were analysed. At each measuring point samples induced with Rifampicin and samples treated with DMSO were analysed. The samples treated only with DMSO were used as negative controls because Rifampicin was dissolved in DMSO.

Two systems biology strategies of gaining new insights were used: the "top-down" approach based on a holistic description leads deductive to new knowledge. This method of obtaining a global view of gene expression based on DNA microarrays was performed using Affymetrix full genome chips allowing the detection of > 20.000 human genes in parallel. In the work presented here the expression of the whole genome was measured for three patients at three points in time (6, 24 and 72 h after induction) in induced- and control samples.

In the scope of this work and the cooperation with the IBVT new data was obtained resulting in a consolidated view on interactions between PXR, HNF4 α and co-activators like SRC-1 and PGC-1.

In parallel an inductive "bottom-up" approach was used. Multiple analyses were performed based on a pre-selection of genes involved in the induction of CYP3A4 with

Rifampicin and the impact of this induction on the metabolic and regulatory networks of the liver cell. This smaller subset of genes could be monitored in more detail as it is possible for the entirety of all genes.

The experimental part of this approach was performed using sub-genome DNA-arrays from Eppendorf and quantitative real-time polymerase chain reaction (qRT-PCR). Especially due to the lower cost, analyses with more measurement points were possible. The dynamic expression profile was measured after 6, 12, 16, 20, 24, 28 and 72 hours of induction using DNA-arrays and 0.5, 1, 4, 8, 12, 16, 24, 30, 36, 42 and 48 hours after induction based on qRT-PCR measurements.

While the "top-down" approach permits the identification of interesting genes, the "bottom-up" method allows the refinement of knowledge starting from previously identified pathways of cellular response.

To analyse gene expression liver cells were harvested at several points in time after induction. The total RNA from induced and control samples were isolated. Depending on the different monitoring methods used, mRNA was transcribed into cDNA or amplified. cDNA synthesis was performed for quantitative detection with qRT-PCR or chip experiments using direct labeling. Alternatively mRNA was amplified and used for chip experiments with indirect labeling (Affymetrix and Eppendorf). To allow detection of nucleic acids on the microarrays, modified bases were incorporated into the newly synthesized nucleic acids. The labeled nucleic acids were hybridized on the microarray and detected using a fluorescence scanner. As the ratio between induced- and not induced samples at different measurement points had to be investigated a "time matched control" design was chosen. This design comprises the parallel hybridization of a control and induced sample for each point in time. Allocating both values allows determination of differential gene expression.

The main results of the work presented here can be summarized as follows: 24 h after induction of primary human hepatocytes with Rifampicin a regulated transcription based on the stimulus was proven. Genes of the xenobiotic metabolism as like several transporters (ATP-binding cassette, sub-family B (MDR/TAP), member 1 (ABCB1) transporter, solute carrier family 1 (glutamate/neutral amino acid transporter), member 4 (SLC1A4) transporter and solute carrier family 7 (cationic amino acid transporter,

system), member 1 (SLC7A1) transporter) as well as several cytochrome P450 enzymes (CYPs) showed increased transcription. A very significant observation was the up-regulation of aminolevulinate delta-synthase 1 (ALAS1). ALAS1 is the regulated gate-keeping enzyme of the heme biosynthetic pathway. Co-regulation of expression of this enzyme with P450 enzymes makes sense as heme is needed for synthesis of the prosthetic group of these monooxygenases. In addition a group of metallothionein proteins was identified to be significantly up-regulated. These enzymes might play a role in detoxification of radical oxygen species originating from P450 action. Furthermore a down-regulation of genes involved in glycogen storage metabolism as glycogen synthase 2 (liver, *gys2*) and phosphoenolpyruvate carboxykinase 1 (soluble, *pck1*) was determined. This was a quite interesting observation as the nutritional status next to the age and sex of the patient has an extensive impact on the whole regulation.

In this context the influence of the induction on lipid metabolism, bile acid metabolism, bilirubin synthesis and degradation was investigated and characterized. All these metabolic functions are located in the liver and are partially carried out by the same metabolic enzymes (CYP3A4) and regulated by the same nuclear receptors (PXR).

Interestingly the main components of regulation including hepatocyte nuclear factor 4, alpha (HNF4 α), respectively the nuclear receptors PXR (nuclear receptor subfamily 1, group I, member 2, NR1I2) and CAR (nuclear receptor subfamily 1, group I, member 3, NR1I3), were shown to be down-regulated. The strong induction of the target genes of these receptors, which was demonstrated in parallel suggested that no increase in expression level seems to be necessary to react on the Rifampicin stimulus. The amount of ubiquitous available receptor protein may be sufficient to allow binding to a high extent at the promoter region of target genes inducing efficient transcription. To verify this hypothesis subsidiary experiments are ongoing to show binding of nuclear receptors to DNA under induced conditions. These binding studies are performed using Biacore assays and ChIP-chip analysis (Hepatosys second phase, Institute for Technical Biochemistry, Stuttgart and Microarray Facility, Tübingen).

The analysis using sub-genome DNA arrays showed considerable individual differences. Especially the induction of cytochrome P450-enzymes (CYPs); glutathione S-transferase A2 (GSA, GSTA2); ATP-binding cassette, sub-family C (CFTR/MRP), member 1 (ABCC1) transporter and cyclin B1 (CCNB1) as well as enoyl-Coenzyme A,

hydratase/3-hydroxyacyl Coenzyme A dehydrogenase (EHHADH) showed remarkable variation of expression in samples from different patients. In almost all cases the change of expression concerning up- or down-regulation was similar but the amount of induction or repression was variable. Nevertheless several similarly induced as well as repressed genes were identified in different patients. It was demonstrated that especially the induction showed a high variation while the down-regulation of genes was in most cases similar in samples from different donors.

Another remarkable observation using these sub-genome measurements was inflammation linked repression of CYP3A4 expression based on interleucin-6 (IL-6).

Induction of CYP3A4 was additionally verified on mRNA level using real-time PCR and on protein level by Western-blot. qRT-PCR assays were further used to analyse a time series experiment, to obtain data for the dynamic behaviour of gene expression of CYP3A4 and PXR. This approach is extendible to other genes, as most of the human genes are commercially available as qRT-PCR assay.

In summary the microarray study performed in the scope of this work gives hints for so far not described regulation mechanisms and could validate many results of former scientific experiments. Next to the comparison of three different commercially available microarray platforms a large number of significant differentially expressed genes after Rifampicin induction could be identified using these platforms. Identified genes play a role in carbon metabolism (gluconeogenesis), in heme biosynthesis, in bilirubin-, bile acid- and lipid metabolism as well as xenobiotic-metabolism and transport. The observed cell response on RNA level over time allowed new insights in the regulation of this complex system. Next to the network reconstruction, based on the full genome array data, the highly time resolved measurements enable dynamic modelling of expression and therefore a mathematical reproduction of the cellular response upon the stimulus. Using the developed models predictions concerning the behaviour of the "system hepatocyte" and the associated regulations will be possible in the future. This model based prognosis will especially allow improved drug development and personalized medicine.

Abbreviations

Abbreviations for gene names used in the work presented here can be found in the appendix or under <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&list>.

°C	degree celsius	Cy3	cyanin-3
μ	10 ⁻⁶ (micro)	Cy5	cyanin-5
μm	micrometer	CYP	cytochrome P450 monooxygenase
μM	micro molar	Δ	delta
A	adenine	d	day
A	ampere	DCC	detection concentration curve
ABC	ATP-binding cassette	ddH₂O	double deionized water
ABCB1	ATP-binding cassette, sub-family B member 1 (also MDR-1, P-gp)	DDT	Affymetrix archive file
Ahr	aryl hydrocarbon receptor	DEPC	diethyl pyrocarbonate
aRNA	anti sense RNA = amplified RNA = cRNA	DMSO	dimethyl sulfoxide
ATP	adenosine triphosphate	DNA	desoxy ribonucleic acid
B	intensity of background	DTT	dithiothreitol
BCIP	5-bromo-4-chloro-3-indolyl- phosphate	e.g.	exempli gratia
bit	binary digit	EC	environmental chemical
β-ME	β-Mercaptoethanol	EDTA	ethylenediaminetetraacetic acid
bp	base pair	ETH	Eidgenössische Technische Hochschule
BSA	bovine serum albumin	EtOH	ethanol
C	cytosine	f	female
CAR	constitutive androstane receptor (NR1I3)	FA	formaldehyde agarose
cDNA	complementary DNA	FAM	6-Carboxy-Fluorescein
cel	Affymetrix data file containing intensities	FCS	fetal bovine serum
CO₂	carbon dioxide	fdr	false discovery rate
COMET	Computer Emulation in Toxicology	FRET	fluorescence resonance energy transfer
COXs	cyclo oxygenases	G	guanine
CSV	comma separated value	g	gram
CTP	cytidine triphosphate	g	gravitation
CV	standard deviation (= Std dev)	GC- RMA	GC-robust multichip analysis
		gfg	gene function gene
		GITC	guanidine isothiocyanate

Gluta Max	stabilized, dipeptide L-alanyl-L-glutamine	MGED	microarray gene expression data
GmbH	Gesellschaft mit beschränkter Haftung	MIAME	minimum information about a microarray experiment
GMO	genetically modified organism	min	minute
GST	glutathione S-transferase	mio	million
GTP	guanosine triphosphate	MM	mismatch
h	hour	mM	millimolar
H₂O	water	MO	MGED ontology
HCC	hepatocell carcinoma	MOAs	monoamine oxidases
HH	human hepatocytes	MOPS	3-(N-morpholino) propanesulfonic acid
HKG	house keeping genes	MPC	Medical Proteom-Centre
IBVT	Institute for Biochemical Engineering	mRNA	messenger ribonucleic acid
IKP	Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology	MRP	multidrug resistance related protein
IS	internal standard	MSD	Mahalanobis squared distance
ITB	Institute of Technical Biochemistry	MT	metallothionein
IU/ml	international units / millilitre	MW	molecular weight
IVT	<i>in vitro</i> transcription	n	10 ⁻⁹ (nano)
JAK	Janus Kinase	n.d.	not determined
k	kilo (10 ³)	NADH	nicotinamide adenine dinucleotide, reduced
kB	kilo bases	NaOAc	sodium acetate
l	litre	NaOH	sodium hydroxide
LB	Luria Bertani	NBT	4-nitroblue-tetrazoliumchloride
LMMA	literature mining and microarray analysing	NCBI	National Center for Biotechnology Information
ln	natural logarithm	NHCS	negative hybridization controls
log	logarithm	nm	nanometer
M	M-value	not tr.	not treated
m	10 ⁻³ (milli)	NR	nuclear receptor
m	meter	NR1I2	nuclear receptor subfamily 1, group I, Member 2 (see PXR)
m	male	NR1I3	nuclear receptor subfamily 1, group I, Member 3 (see CAR)
mA	milliampere	NTCP	sodium dependent taurocholate co-transporting polypeptide
MAGE	microarray gene expression	NTP	nucleoside triphosphate
MAO	monoamine oxidase	OATP	organic anion transporting polypeptide
MAP	mitogen-activated protein	OD	optical density
MAQC	microarray quality control	p	see p-value
MAS	microarray suite		
MDR-1	multidrug resistance 1 (ABCB1, P-gp)		

PB	Phenobarbital	SAPE	streptavidin-phycoerythrin
PCR	polymerase chain reaction	SDS	sodium dodecyl sulfate
Pen/ Strep	Penicillin/Streptomycin	siRNA	small interfering RNA
PGG₂	hydroperoxy endoperoxide prostaglandin G ₂	SNP	single nucleotide polymorphism
PGH₂	prostaglandin H ₂	SOP	standard operating procedure
P-gp	p-glycoprotein (also ABCB1 and MDR-1)	STAT	signal transducers and activators of transcription
PI3K/ Akt	phosphatidylinositol 3-kinase and Akt kinase	SULT	sulfotransferase
PM	perfect match	T	thymine
PMT	photomultiplier tube	t-test	t-test statistic
p-value	the probability of obtaining a result at least as extreme as a given data point, under the null hypothesis	t	time
PVDF	polyvinylidene difluoride	TAE	tris-acetate buffer (pH 8.0) with EDTA
PXR	pregnane X receptor (NR1I2)	TAMRA	6-Carboxy-tetramethyl- rhodamine
PZN	Pharmazentralnummer (central number for pharmaceuticals)	TBS	tris buffered saline
qRT PCR	real-time polymerase chain reaction	TCA	tricarboxylic acid
q-value	measures the minimum FDR that occurs when calling the test significant	TEMED	tetramethylethylenediamine
RefSeq	NCBI Reference Sequence	TF	transcription factor
rif	Rifampicin	TfbI	transformation buffer I
RIFA	Rifampicin	TfbII	transformation buffer II
RIN	RNA integrity number	tiff	tagged image file format
RLT	guanidine isothiocyanate (GITC) containing buffer	TRIS	tris(hydroxymethyl)aminomethane
RMA	robust multichip analysis	TTBS	Tween-TBS-buffer
RNA	ribonucleic acid	TTP	thymidine triphosphate
RPA	ribonuclease protection assay	txt	text
rpm	rounds per minute	U/ml	units per millilitre
RT	reverse transcription	UGT	UDP Glucuronosyltransferase
RXR	retinoid X receptor	UNG	uracil-DNA glycosylase
s	second	UTP	uridine triphosphate
s	standard deviation	UV	ultraviolet
S	Svedberg constant	UV/vis	ultraviolet-visible
Σ	sum	V	volt
S	intensity of spots	V.	version
SAGE	serial analysis of gene expression	v/v	volume per volume
		w/v	weight per volume
		WME	Williams Medium E
		XML	extensible markup language
		θ	temperature

1 Introduction

This introduction is intended to provide a conceptual, experimental and methodological foundation on microarray studies performed in a systems biology context like in the work presented here. In Chapter 1.1 an introduction to the growing field of systems biology is given. The difference to classical approaches is emphasized and the benefit of investigating biological questions in a holistic manner is shown. In Chapter 1.2 the framework of this project - Hepatosys - is introduced. The basic research ideas of investigating the liver metabolism and regulation in hepatocyte cells are explained. A more detailed view on these liver specific functions presents Chapter 1.3 which deals with the xenobiotic metabolism. Chapter 1.4 cover basic concepts of microarray technology. Separate chapters are devoted to development of microarray technology, application and different kind of array formats, planning of experiments and experimental design, technical aspects of microarray experiments, standardization and finally microarray data analysis. In Chapter 1.5 the objectives, of the work presented here, are summarized.

1.1 Systems Biology - a New Approach

What does systems biology mean? Imagine the following situation: a person feels sick, high fever, the muscles are aching and he feels miserable: diagnosis flu. His physician prescribes antibiotics and few days later the patient feels much better. What exactly happens at the molecular level when the person becomes sick having the flu takes an antibiotic and gets rid of both the flu and the drug has been a mystery to science. Additional other questions remain, e.g. why people heal certain cancers while others don't, and why some people never get sick at all.

Dependencies and relationships leading to such observations like the healing of a patient are what biologists investigate nowadays and what is called systems biology. The study on how systems interact in a complex biological context (like the highly

cross-linked organ liver in the work presented here) is one of the challenging scientific question now and will continue to be in the future.

The systems approach will revolutionize biology in the 21st century and beyond. It is a new way of thinking about biology and will make better, more personalized treatment of diseases possible.

Biology in the classical point of view is the study of life. It is a field of empirical sciences that examines the structure, function, growth, origin, evolution, and distribution of living objects. It classifies and describes the various forms of organisms, how organisms function, how species come into existence, and the interactions they have with each other and with the natural environment.

There were already system-like approaches in traditional biology like the predator-prey relationship or the investigation of ecosystems. However new technical developments were the prerequisite for systems biology as we understand it today. Till the 21st century, studies, especially at the molecular level, focused on the analysis of single components of whole systems, like genes, proteins or cells. Several years ago it was technically not feasible to investigate hundreds or thousands of compounds in parallel. Therefore it was necessary to reduce complexity, by investigating all singular components of the system (e.g. an organism) individually. The hope was to be finally able to understand their interactions and dependencies by putting all parts together. The investigated components had to be reassembled in a bottom-up like approach to derive the behaviour of the complex system from the activities of multiple singular compounds. For some aspects in biological-, biochemical-, immunological- or medical research this strategy worked well. The described approach has successfully identified most of the today known components, many of the interactions and basic principles in metabolism, gene regulation, cell cycle, immune response and so on. Unfortunately it offers no convincing concept or method to understand how system properties emerge. The total system often has characteristics which can't be identified by putting together the single components. The multidimensional causes and effects in biological networks are better addressed by observing quantitative measures of multiple components simultaneously and by rigorous data integration with mathematical models (Kitano, 2002; Sauer *et al.*, 2007).

The term "systems biology" appeared the first time in the 1940s when the biologist and philosopher Ludwig von Bertalanffy (1901-1972) developed a kinetic theory of stationary open systems and his "General System Theory". He was one of the first, who applied the system methodology to psychology and the social sciences. He extended concepts of a holistic analysis of biological systems to a more general idea of "developing a theory of systems" fighting against the fragmentation of knowledge in science (von Bertalanffy, 1951).

"Systems biology", describing molecular interactions in higher order of complex systems, started in 1952 when the British neurophysiologists and Nobel Prize winners Hodgkin and Huxley developed a mathematical model for the action potential along the axon of a neuronal cell. In 1960, Denis Noble presented a computer model of a beating heart. In the 1960s and 1970s the systems biology began to study complex molecular systems like the regulation of metabolism but was often limited by experimental and computational power.

The success of molecular biology in the 1980s led to a minor interest in quantitative modelling of biological processes in a larger scale. Experimental approaches on single gene level (deterministic genetic view) looked more promising. However a lot of aspects were still unknown, couldn't be explained and required more sophisticated approaches to be unravelled. But technology has caught up with the required analytical power to investigate very complex systems. The publication of the whole *Haemophilus influenzae* genome in 1995 was the beginning of a change in biological research. The start of whole genome sequencing and other high throughput experimental technologies transformed biological research from a relatively data poor discipline into one, where huge data sets are quite common. Following the whole genome sequencing, technical innovation provided genome-scale measurements for almost all molecular species in a cell (Joyce and Palsson, 2006). With the start of functional genomics in the 1990s, large quantities of good quality data were available and allowed a closer look for the pluralism of causes and effects in biological networks. Various technologies generate genome-scale, or so-called "omics" data which provides systems-level measurements for almost all types of components in an organism. Especially the high throughput analysis of mRNA (transcriptomics), protein (proteomics) and metabolites (metabolomics) are promising tools to investigate the systems in a holistic way. Also

the computing power was no longer a limitation, allowing the implementation of larger and more realistic models. Meanwhile the amount of data in public databases, especially for *Homo sapiens*, has tremendously increased (Figure 1) and offers systems biologists a great opportunity to work on.

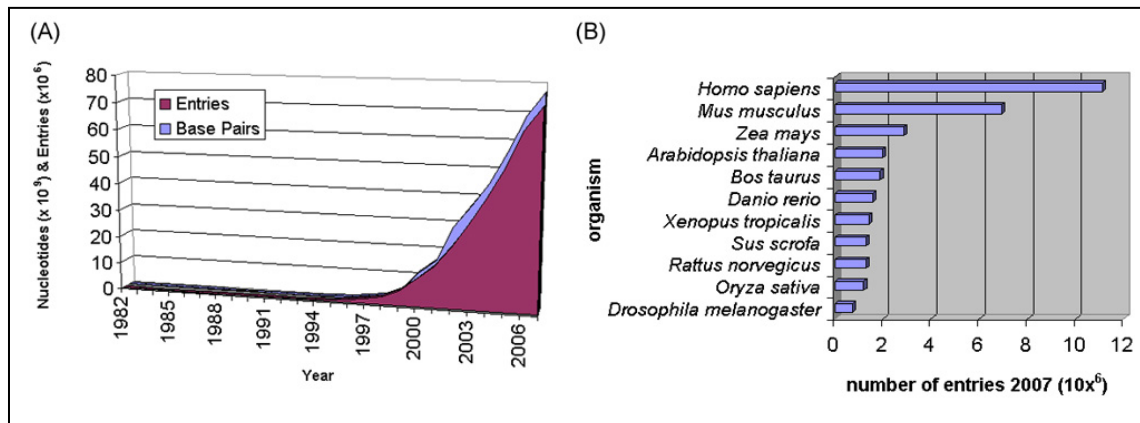


Figure 1: Growth of public Gene bank databases. (A) The number of nucleotides in billions of bases (10^9) and entries (10^6) submitted to the Genbank database from 1982 to 2006 and (B) the eleven organisms that received the greatest number of GenBank submissions for July 2007 (Data courtesy of NCBI).

With the beginning of the 21st century institutes of systems biology were established in Seattle and Tokyo. Several research institutes dedicated to systems biology (like the ETH Zurich or the University of Stuttgart) emerged and a working to accomplish the high expectations in this area.

Beside the technical progress there is another important point which makes systems biology different from previous approaches. What it makes so powerful is the way of integration and combination of several basic research as well as applied research areas. The interplay of multilateral scientific disciplines is an absolute prerequisite for the holistic approach of systems biology. Close interaction between the generation of experimental data and theory based computational modelling and simulation is required in any case. Systems biology or computational biology, as it is often referred to in English literature, is an interdisciplinary field that applies the techniques from biology, chemistry, physics, computer science, applied mathematics and statistics to study and model the function and interactions between components of biological systems.

Finally the question how systems biology works has to be discussed. An overview of the used strategy is given in Figure 2.

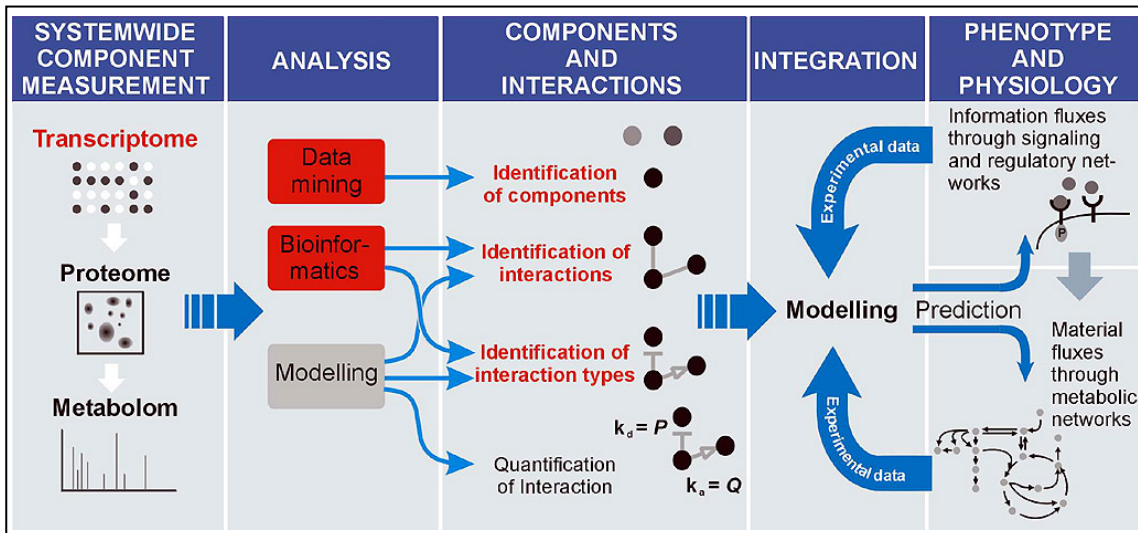


Figure 2: Overview of Systems biology workflow based on Sauer et al., 2007. Data from "omics" technologies provide the basics for identifying and quantifying interactions and thus allow to build metabolic and regulatory networks. These models give predicting power and the ability to give recommendations for further experiments. These experiments will improve the models, starting the cycle of systems biology. The parts shown in red were done in this work.

In a first step, "system wide component measurement", the relevant parts and components of the system have to be identified. To collect gene expression data using DNA microarrays was the major task of the work presented here. Taking into account the huge amount of cellular components, reconstruction and definition of relevant targets is quite challenging. In general two strategies for this purpose are applied: the bottom-up and top-down approach. Using the bottom-up approach to reconstruct a network, already known biological knowledge about single components and their interactions is combined to build up the system or at least parts of it.

In contrast to start with single measurements the top-down approach uses holistic data sets ("omics" data) to reconstruct a model of the whole system. Using mathematical algorithms and optimization strategies, topologic as well as kinetic characteristics of the network can be predicted and integrated into the computational models. This approach is named "reverse engineering".

Both approaches have their advantages and disadvantages. The bottom-up approach suffers from the often missing knowledge about the necessary components for the network. It is a challenge to consider all important variables during the experimental measurements and reconstruction of the system. Often the lack of knowledge still makes it impossible to elucidate the relevant components and their interactions.

The top-down approach however needs enormous computational power and the reconstruction could fail due to the high number of combinatorial possibilities. In the field of top-down approaches and reverse engineering a lot of research is ongoing and new methods and tools are developed to overcome the problems related to these multi-dimensional data sets. Both strategies can be seen as complementary approaches.

In this work both ways were combined to analyse the regulation, signal transduction and metabolism in the human liver. Primary human hepatocytes were treated with Rifampicin and the resulting changes in gene transcription were analysed. On one hand a small subset microarray from the company Eppendorf was used to investigate the expression of 151 genes in more detail (more points in time with smaller time slots). This analysis allowed the monitoring of the dynamic behaviour of the system. On the other hand full genome microarrays from Affymetrix were used for the top-down approach. These arrays measure the expression of all human genes in parallel allowing a holistic view on the change in expression of the whole system. An overview of these two basic strategies is given in Figure 3.

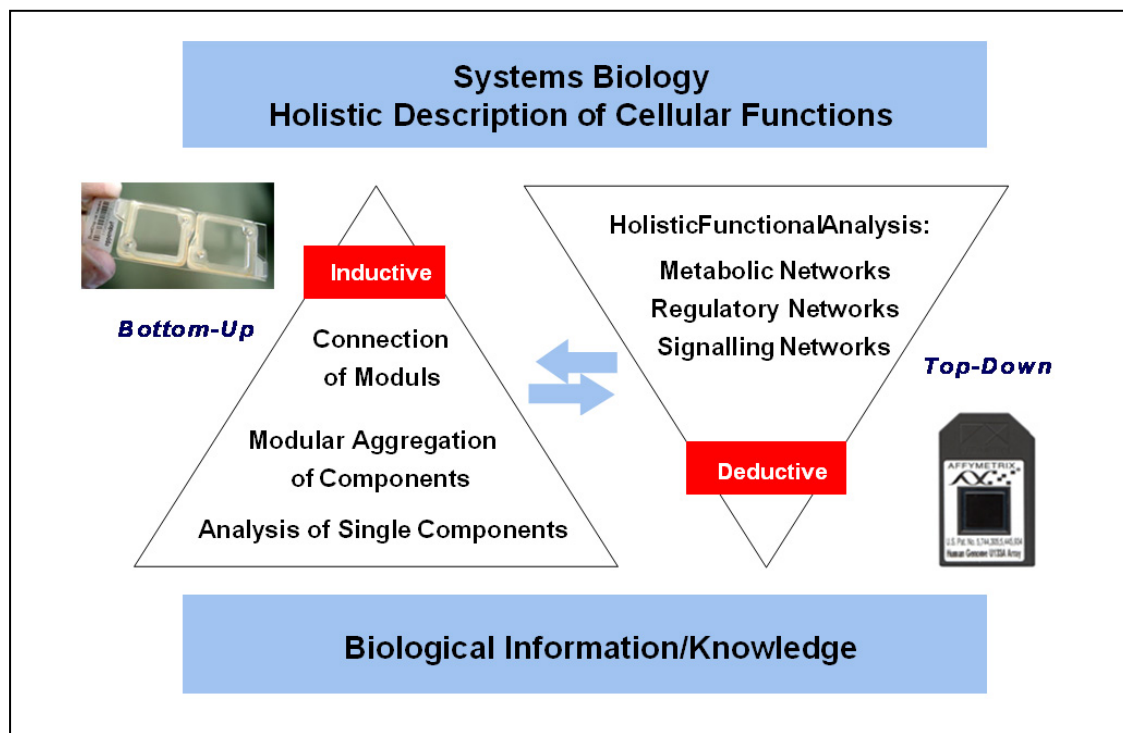


Figure 3: Bottom-up and top-down approaches in systems biology. Using available biological knowledge single components of the whole system can be analysed in a bottom-up approach using subset arrays containing 151 genes. In addition all human genes can be analysed using a full genome array and the networks are reverse engineered using a top-down approach. Both methods gain new biological knowledge and insights in the system behaviour of the cell (graph: based on Mauch *et al.*, 2005).

After reconstruction of the topology of metabolic- and signal transduction networks the model structure has to be filled with kinetic data to predict system behaviour and the two networks have to be integrated. This integration of signalling- and metabolic networks is one of the major challenges in systems biology. However it is indispensable because phenotypic behaviour strongly depends on signal transduction and the distribution of metabolic fluxes.

1.2 "Hepatosys" - a Computational Approach to Understand Functional Aspects of the Liver

The liver has many functions. Some of these functions are: to break down fats, produce urea, make certain amino acids, filter harmful substances from the blood, storage of vitamins and minerals and convert glucose to glycogen and therefore maintain a proper level of glucose in the blood. The liver is also responsible for producing cholesterol. To deal with all these tasks, the liver is a very complex biochemical factory. It has to modify and transform over 10,000 substances each day to fulfil the tasks of providing the body with essential nutritional building blocks like sugars, lipids and proteins and to get rid of toxic metabolites like drugs and toxins (detoxification). All these reactions take place in the well structured organ which is built of several circuits. First blood is carried to the liver via two large vessels called the hepatic artery and the portal vein. The hepatic artery carries oxygen-rich blood from the aorta. The portal vein carries blood containing digested food from the small intestine. These blood vessels subdivide in the liver repeatedly, terminating in very small capillaries. Each capillary leads to a lobule. Liver tissue is composed of thousands of lobules, and each lobule is made of hepatic cells, the basic metabolic cells of the liver. These hepatocytes are the research object in this study as they are the key player of a versatile and complex metabolic- and regulatory system. The reason for choosing the liver as model system is based on the manifold tasks of this organ and the resulting applications in medicine, pharmaceutical industry and food industry. The goal, an *in silico* vitreous liver cell would allow to reduce time and costs for developing new drugs for example. In addition the number of experiments with animals can be reduced.

The use of systemic genomic, proteomic and metabolomic data to construct models of complex biological systems and diseases is known as systems biology (Ideker, 2004). The approach by integrating knowledge from diverse datasets (transcriptome, proteome, metabolome, interactome) to iteratively refine our knowledge about the liver is also represented in the sub-network structure of our project part ("COMET") of "Hepatosys". "COMET" is an abbreviation for "COMputer Emulation in Toxicology" and represents the final goal of this project. This project part is especially interested in all mechanisms involved in detoxification.

Various datasets were generated by the MPC (Medical Proteom-Centre) in Bochum, dealing with proteomics, the IKP (Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology) in Stuttgart working on the metabolome and kinetic measurements as well as the work presented here regarding the transcriptome done at the ITB (Institute of Technical Biochemistry) in Stuttgart. Based on this a representative model of the liver in concern to metabolic- as well as regulatory- function is generated in the modelling group at the IBVT (Institute of Biochemical Engineering).

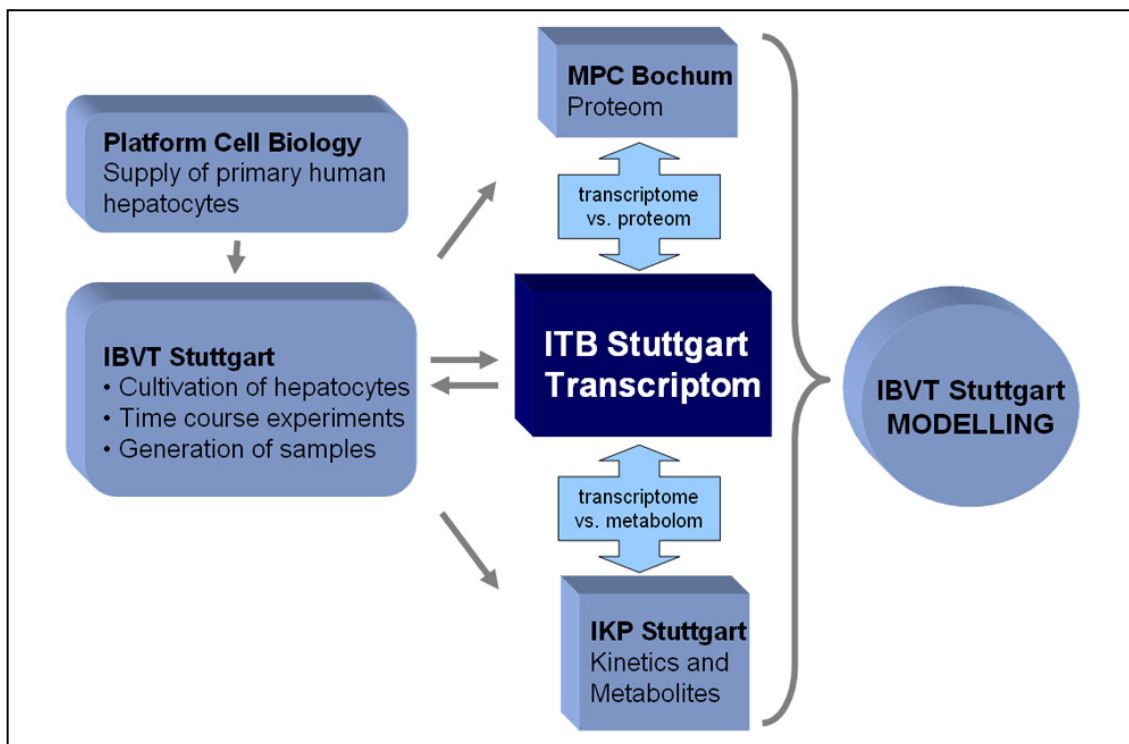


Figure 4: Integration of transcriptome measurements using DNA microarrays in the cross-disciplinary project part detoxification in the whole "Hepatosys" network. Human hepatocyte cells are delivered by the platform cell biology and cultivated at the IBVT (Institute of Biochemical Engineering). The time/dose experiments are performed in collaboration with the ITB (Institute of Technical Biochemistry). The cell culture lab at the IBVT distributes the cells to the sub-projects analyzing proteom, transcriptome and metabolom. The data of the different groups is collected and send to the IBVT where the modelling groups try to integrate it in their quantitative models.

In Figure 5 a graphical representation of the groups and fields studied in "Hepatosys" is given. The coordination of the different parts of detoxification is based on the gene regulation shown here in red. All transport processes and metabolic function are tightly regulated to work together. The investigation of regulation is part of the work presented here as well as two other groups (JAK-STAT and MAP-Kinase group). While these two groups are focused on specific, known fields of regulation, this work presented here is

interested in a holistic analysis of the regulation in human hepatocytes to gain new insights. The regulation data is further enhanced with metabolic and proteome data from the previously mentioned groups of "COMET" as well as other groups in the "Hepatosys" network shown here in blue.

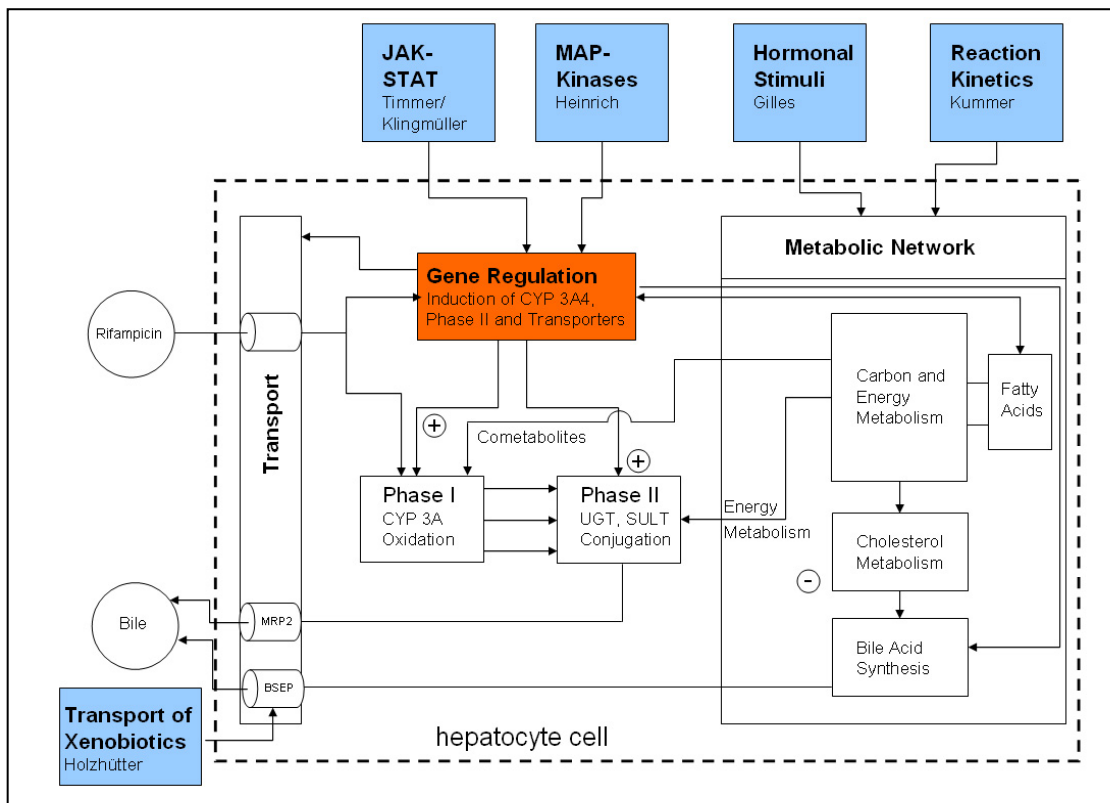


Figure 5: Graphical representation of groups working in the "Hepatosys Network". A central part in this network (shown here in orange) is the gene regulation which is investigated by DNA microarrays in the work presented here. In blue connecting nodes to other groups of the "Hepatosys Network" are shown. The groups and their research objective are mapped over a functional flow-chart of a hepatocyte cell. (graph: Insitute of Biochemical Engineering, modified by Thomas Reichart)

The project was positively reviewed by an international evaluation committee because it combines results from different groups working on special aspects of the liver (like regeneration, regulation, detoxification, and others). The objective to get an overall view is unique in the world and the scientific program is of high quality and gets high attention.

1.3 Xenobiotic Metabolism

The central role of the liver is the detoxification of foreign (xenobiotic) compounds. Every eukaryotic organism is exposed to a wide array of xenobiotic substances from food components to environmental toxins to pharmaceuticals which might cause toxic effects to the body. To avoid these undesired effects of substances or their metabolites the organisms developed complex enzymatic systems to convert these compounds and get rid of them. This whole process is called detoxification and involves a huge number of enzymes and pathways. An overview of several well investigated and important pathways is given in Figure 6.

The systems are highly complex, show a great amount of individual variability and are extremely responsive to an individual's environment, lifestyle and genetic background which make the elucidation of regulatory and metabolic principles quite challenging. The hypothesis that xenobiotics are transformed to water soluble substances and excreted through the urine was posted in the 18th century. In 1842 Keller proofed this theory in a self experiment. He ingested benzoic acid and could detect hippuric acid in his urine, which is formed after conjugation of glycine with benzoic acid. Research continued for more than 100 years by identifying several metabolites and conjugation reactions. Nowadays it is clear that during this detoxification the hydrophobic substances become more polar which allows a more readily excretion from the body via bile, faeces and urine. Furthermore they are unlikely to have access to hydrophobic, membrane bound receptors, avoiding undesired activation of e.g. signal transduction pathways.

Proteins catalyzing the biotransformation of xenobiotics are widely distributed in the body. They can be found e.g. on the first barrier between environment and body, the skin, and of course the intestine which is highly exposed to xenobiotics due to the permanent uptake of food. A considerable part of detoxification occurs in the liver as all the blood from the gastrointestinal tract containing a lot of foreign substances is flowing through the liver. So the liver is the second barrier for xenobiotic substances which were resorbed by the stomach and small intestine before they reach the systemic blood stream and affect the whole body.

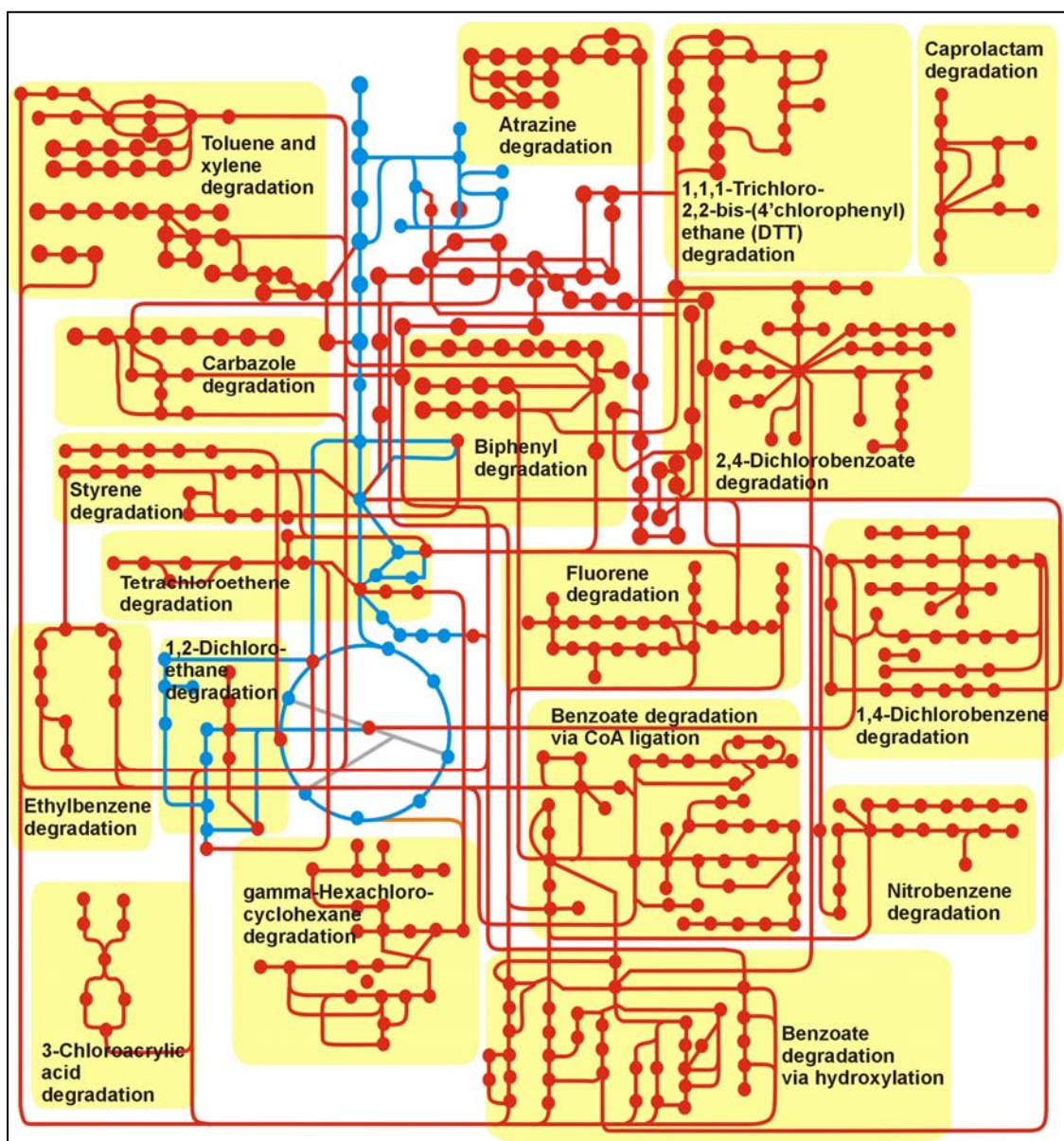


Figure 6: Well investigated and important parts of the xenobiotic metabolism. This biodegradation overview is based on KEGG pathway map 01196 representing molecular interaction networks of xenobiotic metabolism. Shown in blue is the central carbon metabolism.

The biotransformation and the associated detoxification of xenobiotics are divided into three parts (Figure 7).

In 1947, Williams introduced the concept of phase I and phase II biotransformation ("Detoxification Mechanisms"). Phase I are the so called "functionalization reactions" and the first enzymatic defense against foreign substances. During these oxidation-, reduction- or hydrolysis reactions, functional groups (OH, SH, NH₂, or CO₂H) are introduced or activated. The most important enzymes of phase I are members of the

cytochrome P450 monooxygenases gene family (CYPs). CYPs use oxygen and NADH as cofactor to add reactive groups to the xenobiotics. At least 10 families of phase I activities have been described in humans. The major P450 enzymes involved in detoxification of xenobiotics as well as endogenous molecules are CYP4A, CYP1A1, CYP1A2, CYP2D6 and the CYP2C.

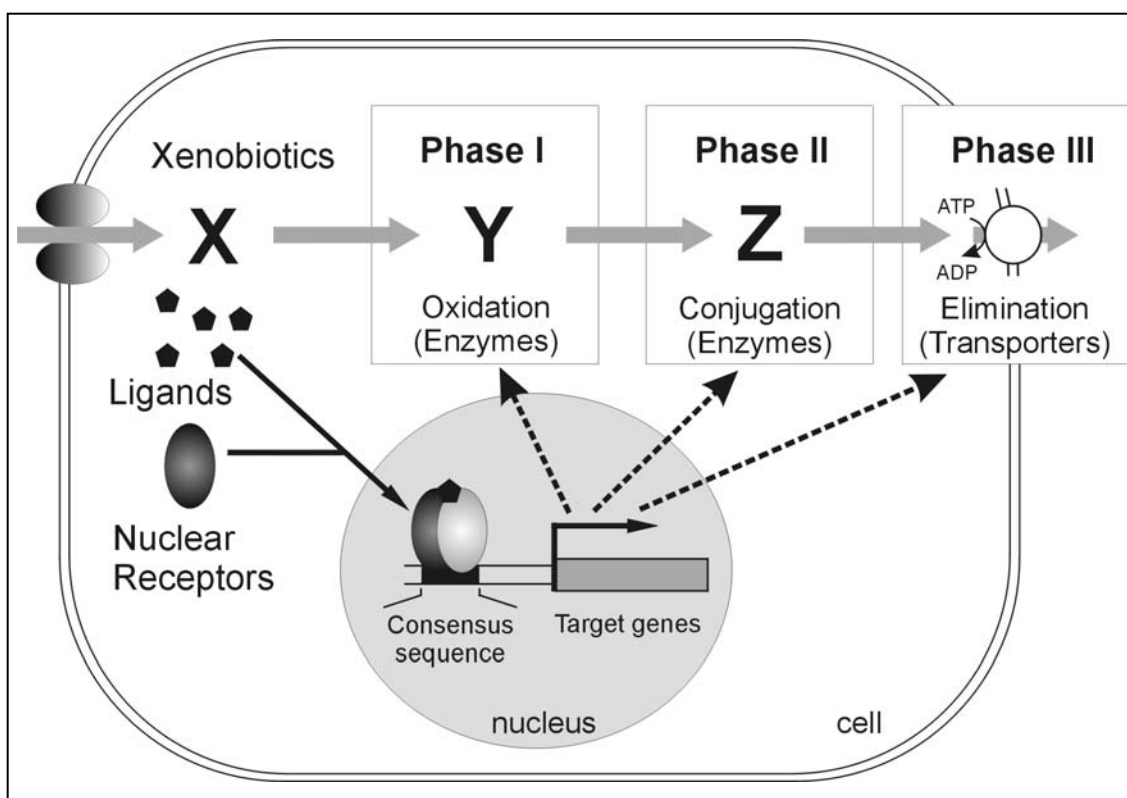


Figure 7: Schematic illustration of phase I, phase II and phase III of the drug metabolizing system. The xenobiotics (X) are activated by phase I enzymes. The product from phase I is converted into a hydrophilic compound (Z) by phase II enzymes and finally eliminated by transporters in phase III. The whole system is regulated by interaction of nuclear receptors and hydrophobic ligands which trigger the expression of the catalytic enzymes. Graph based on Nakata et al., 2006.

Next to this family of monooxygenases there are some other important phase I enzymes. Monoamine oxidases (MOAs) for example play an important role for the oxidative deamination of monoamines ingested in food. Oxygen is used to remove an amine group from a molecule, resulting in the corresponding aldehyde and ammonia. This reaction is also important for the inactivation of dopamine a natural neurotransmitter. Cyclo oxygenases (COXs) are enzymes responsible for formation of important biological mediators called prostanoids like prostaglandins, prostacyclin and thromboxane. Pharmacological inhibition of COX can remove the symptoms of

inflammation and pain which is used in well-known drugs as aspirin and ibuprofen. COX converts arachidonic acid to prostaglandin H₂ (PGH₂). The enzyme contains two active sites: a cyclooxygenase site, where arachidonic acid is converted into the hydroperoxy endoperoxide prostaglandin G₂ (PGG₂) and a heme with peroxidase activity, responsible for the reduction of PGG₂ to PGH₂. The reaction proceeds through H atom abstraction from arachidonic acid by a tyrosine radical generated by the peroxidase active site. Two O₂ molecules then react with the arachidonic acid radical, yielding PGG₂. The hydroperoxidase activity is also involved in detoxification reactions, as it can utilize a variety of compounds as electron donors, including drugs such as phenylbutazone and sulindac. Other xenobiotics such as benzo[*a*]pyrene can also undergo co-oxygenation during hydroperoxidase reactions (DeWitt and Smith, 1988). The three examples mentioned here show, that there is a strong interaction between detoxification of xenobiotic substances and the endogenous metabolism. The interaction between environmental chemicals (ECs) and endogenous metabolites was recently described (Hodgson and Rose, 2007). A clear separation of both parts of metabolism is not possible and makes interpretation of obtained results quite challenging.

As a consequence of this first detoxification step, molecules are formed which are more reactive than their substrate. If these activated intermediates are not further metabolized by phase II reactions they might damage DNA, RNA and proteins. Therefore in phase II, enzymes like glucuronosyltransferases, sulfotransferases, transacylases and acetyltransferases conjugate hydrophilic, endogenous substances to the functionalized products from phase one reactions. Examples of these conjugation reactions are the formation of *O*- and *N*-glucuronides, sulfate esters, various α -carboxyamides and glutathione adducts, all with increased polarity relative to the unconjugated molecules. The contribution of phase II to the detoxification has received less attention in academic research and clinical practice and is therefore less precisely understood.

Elimination of xenobiotics is done by transport processes, catalyzed by uptake transporters as well as ATP-dependent export pumps in the phase III. The uptake of substances from the blood stream to hepatocyte cells is achieved by sodium dependent taurocholate co-transporting polypeptide (NTCP) and the family of organic anion transporting polypeptides (OATPs). Efflux transporters are e.g. ATP-dependent P-

glycoprotein (P-gp, or MDR1, ABCB1 = ATP-binding cassette sub-family B member 1) the product of the multidrug resistance gene (MDR1) and the multidrug resistance-associated proteins 2 and 3 (MRP2 and MRP3). Up to now more than 40 different human ATP-binding cassette (ABC) transporter genes have been discovered and some of them e.g. ABCB1, ABCB11, ABCC1, ABCC2 and ABCG2 have been demonstrated to be involved in the transport of xenobiotics and metabolites. Antiporter expression in the intestine is co-regulated with intestinal phase I CYPs indicating the supporting role of transporters to the detoxification process. A schematic representation of these functionalization and detoxification reactions is given in Figure 8.

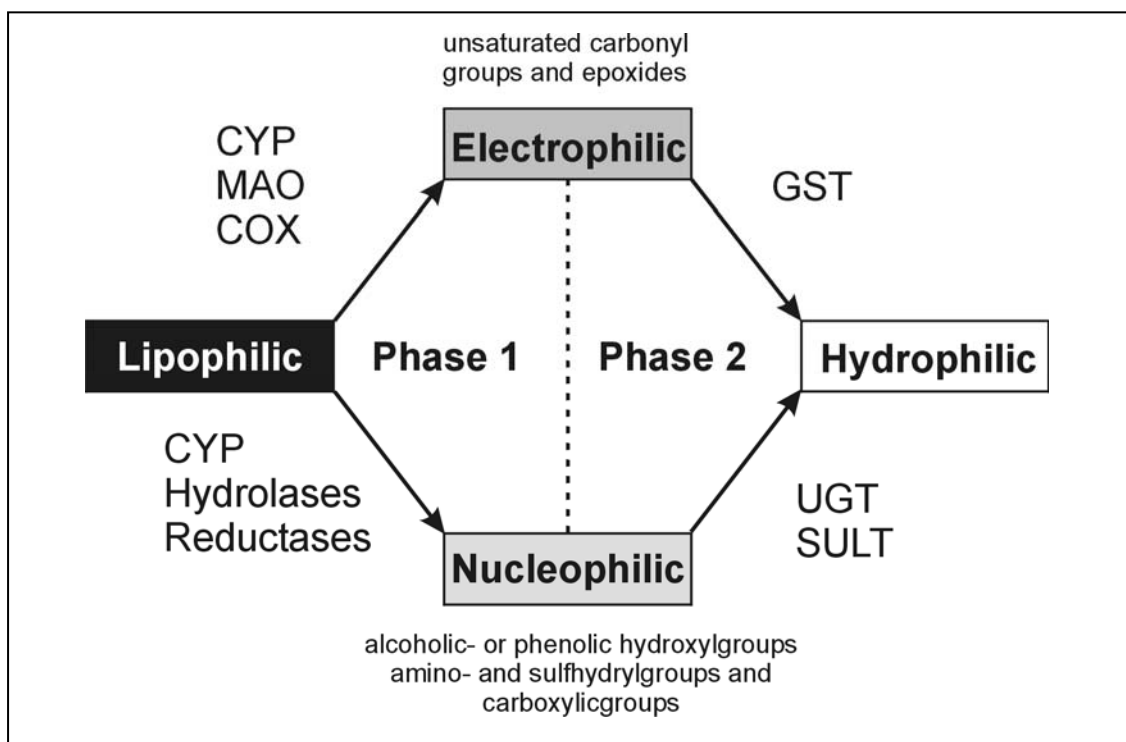


Figure 8: Schematic overview of the xenobiotic metabolism. Lipophilic substances are activated by CYP, MAO, COX, hydrolases and reductases in a so called phase 1 reaction and then transformed in the phase 2 reactions catalyzed by GST, UGT and SULT into hydrophilic substances which can be excreted.

Studies on the genetic regulation of drug metabolizing enzymes and drug transporters are of great interest to understand the molecular mechanisms of drug response and toxic events (Nakata *et al.*, 2006). The transcriptional regulation of catalytic enzymes of phase I, II and III is mainly done by nuclear receptors (NR) and hydrophobic ligands.

These NRs build a family of ligand-activated transcription factors (TFs). In some cases dimerization and binding of a ligand leads to an activation of NRs, that triggers the expression of their target genes. These target genes include metabolizing enzymes such as cytochrome P450 (CYPs) or transporters. The most important NRs for the xenobiotic response system are aryl hydrocarbon receptor (Ahr), pregnane X receptor (PXR) retinoid X receptor (RXR) and constitutive androstane receptor (CAR). As Ahr, PXR and CAR can interact and consequently be activated by intracellular xenobiotics, they are often called "xenosensors" (Kliewer, 2003; Pascussi *et al.*, 2004).

The challenge to understand detoxification continues and new approaches like holistic system oriented experiments are promising attempts to get a better understanding on how the body can handle such a wide range of unknown compounds. The collection of different enzymes with broad substance specificity which evolved for this purpose and their sophisticated regulation are target of our investigation.

1.4 Microarray Technology

Quantitative monitoring of gene expression on printed glass slides originated in the Brown group in the mid 1990s (Schena *et al.*, 1995). The advantage of this new technology compared to the 1991 to 1993 developed real-time PCR and the classical northern blotting is the high capacity. It allows the analysis of expression levels of thousands of genes in parallel. Microarray analysis is an outstanding technology in biological sciences as it involves high technological complexity, combined expertise from many different disciplines such as biology, chemistry, physics, engineering, mathematics and computer science and provides a semi-quantitative and systematic view on biological systems.

The following section gives a summary of the most important discoveries and breakthroughs in the field of biology which led to the development of microarray technology.

In 1949 Pauling and co-workers discovered the coherence between genetic mutation and disease while investigating haemoglobin from sickle cell patients. They concluded that the changes in the haemoglobin gene were responsible for the different protein in sickle cell patients. This remarkable publication in *Science* (Pauling *et al.*, 1949) paved the way for the molecular genetics analysis of diseases and genetic aberrations. At that time no one was thinking of DNA microarrays. However these studies indicated a conceptual proof for the use of microarrays in genetic screening, testing and diagnosis.

With the discovery of Watson and Crick (1953) who published the chemical structure of DNA in their remarkable letter in *Nature* (Watson and Crick, 1953) the development of DNA microarrays got its foundation. They suggested specific base pairing interactions (A-T and G-C) which are a fundamental recognition step during the hybridization on DNA microarrays. For the discovery of the double helix, which was one of the most important findings in twentieth-century science, they were rewarded with the noble price in 1962 (<http://nobelprize.org>). At about the same time Kornberg and Ochoa discovered DNA and RNA polymerase enzymes which link single nucleotides together into DNA and RNA chains. In 1959 they got the noble price for this discovery of "biological synthesis of ribonucleic acid and deoxyribonucleic acid"

(<http://nobelprize.org>). Polymerases are one of the most commonly used enzymes in the laboratory not only in microarray experiments but also for recombinant DNA technology like e.g. polymerase chain reaction (PCR). Another breakthrough which changed a fundamental dogma of biology was achieved by Baltimore, Temin and Mizutani who discovered a special DNA polymerase known as reverse transcriptase. The flow of genetic information which was thought to be from DNA to RNA and not the other way round was reversed by their findings. In 1972 Baltimore, Temin and Dulbecco received the noble price for their investigations on the interaction of (RNA) viruses and DNA (<http://nobelprize.org>). The RT reaction is again one important step in reverse transcribing and labeling of mRNA targets for microarray analysis and has many other practical applications in life science.

Edwin Southern was the first researcher who described the use of labeled nucleic acids for hybridization of DNA molecules which were immobilized on a solid support (Southern, 1975). With the so called Southern blot it was possible to detect one gene, but so far only one. During this time a lot of nitrocellulose and nylon filter methods were developed, creating the basis for the development of microarrays 20 years later, allowing for monitoring and analysing thousands of genes in parallel.

The first DNA array was described by Grunstein and Hogness (Grunstein and Hogness, 1975). The two scientists used a nitrocellulose based array of bacterial colonies to isolate fruit fly genes. Davis and co-workers identified the first differentially expressed genes from higher organisms (Benton and Davis, 1977). At that time Maxam and Gilbert as well as Sanger and co-workers developed DNA sequencing technology and received 1980s Nobel Prize for "their contribution in determination of base sequences in nucleic acids" (<http://nobelprize.org>). The Sanger method was used to sequence the human genome (Venter *et al.*, 2001) and has provided much of the sequence information stored in databases building the foundations of manufacturing DNA microarrays. One of the most important developments in modern biochemistry was also rewarded 1980 with the Nobel Prize in chemistry. Paul Berg received it "for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA" (<http://nobelprize.org>). The use of recombinant DNA technology offered a lot of new possibilities and allowed e.g. preparation of cloned libraries for microarray manufacturing and analysis. An extraordinary technical development

changing biotechnology was the polymerase chain reaction (PCR), nowadays an indispensable tool in molecular biology. In the 1980s Mullis and co-workers established this technique to generate millions of copies of DNA from a small amount of starting material. PCR was a technical breakthrough and Mullis received the Nobel Prize in 1993 "for his invention of the polymerase chain reaction (PCR) method". Today PCR is widely used in cloning experiments, in microarray manufacture and diagnostic applications.

Finally the use of fluorescent dyes for labeling of biological samples prepared the ground for the fluorescent labeling and detection techniques used in microarray analysis. And last but not least Hans Lehrach and co-workers pioneered in the use of robots for high throughput arraying genomic DNA clones on a nylon filter (Nizetic *et al.*, 1991). Despite the large format this pioneering work was quite close to what is still used today.

Concerning the proceedings in assay development the goal was to improve the number of measurements which could be analysed in parallel.

One of the new emerging techniques like the ribonuclease protection assay (RPA) a method for detecting, quantifying, and mapping of specific mRNA transcripts (Zinn *et al.*, 1983; Melton *et al.*, 1984) allowed at least a slight multiplexing. The method is an adaptation of the S1 nuclease assay where RNA-, instead of DNA probes are used, and single stranded specific ribonucleases replace S1 nucleases. Treatment of an RNA:RNA duplex with ribonuclease is more reproducible than treatment of a RNA:DNA duplex with S1 nuclease. However, both S1 assays and RPAs can be used to analyse mRNA transcript abundance and structure. Such kind of solution hybridization assays provide greater sensitivity than hybridization protocols that rely on RNA bound to a solid support (e.g., Northern blots, dot blots). Multiplex reactions (multi probe RPA) with about 10 probes are possible as long as the protected fragment sizes are significantly different. Another promising method called serial analysis of gene expression (SAGE) emerged in 1995 (Velculescu *et al.*, 1995; Saha *et al.*, 1997).

SAGE provides a statistical description of the mRNA population present in a cell without prior selection of the genes to be studied. First of all mRNA is isolated and used with a biotin-labeled oligo-dT primer conjugated with a magnetic bead as template for cDNA synthesis. The cDNA is digested with an "anchoring" restriction enzyme

creating a 3'-overhang. The DNA is purified using the magnetic field. The sticky 5' overhangs are used to ligate a recognition sequence for a "tagging" restriction enzyme. After digestion with this second restriction enzyme which creates a staggered cut with an overhang of 10 bases this overhang is filled blunt end creating a unique 14 to 15 bp sequence (including anchoring restriction site), the so called "SAGE tag" which is specific to each gene. The abundance of each tag is proportional to the amount of mRNA in the original RNA sample. Two tags are then tail to tail ligated, amplified by PCR and cloned creating a series of randomly organized bitags (20 - 24 bp) each flanked with the sequence of the anchoring enzyme. Finally sequencing reveals the abundance by counting the tag frequency.

SAGE has become one of the functional genomics methodologies used in many groups in academia and industry for novel gene and pathway discovery, for biomarker identification, and for gene expression profiling of numerous diseases and normal conditions in multiple species.

Starting in the 1990s major technological advances enabled scientists to complete the sequences of a variety of organisms including viruses, bacteria, invertebrates and finally in the full (draft) sequence of the human genome (Lander *et al.*, 2001). This tremendous increase in whole genome data increased the demand for high throughput methods which would allow parallel measurements of thousands of transcripts in parallel and therefore permitting the translation of genomic sequence information into functional biological mechanisms.

In 1995 Mark Schena, the "father" of microarrays developed a fluorescence based microarray (Schena *et al.*, 1995) and set off an avalanche of new developments and investigations in this area. A timeline of microarray history is presented in Figure 9.

The first microarray experiments were performed using cDNA as immobilized probe and this kind of array is still the most commonly used (about 65 %). A cDNA is a nucleic acid molecule derived from mRNA via reverse transcription reaction and comprises 500 - 2500 base pairs. The use of PCR products (cDNA) for hybridization studies has several advantages. No gene sequence information is required as the cDNAs can directly be amplified from a gene library of interest. After the array experiment the genes identified as important can be sequenced. In addition their large size provides

high potential for hybridization, yielding in strong fluorescent signals. Finally PCR is a commonly used technique that is easily implemented in a standard laboratory compared to more expensive and complicated technologies, such as oligonucleotide synthesis. A disadvantage of PCR generated targets is that their relative large size provides extensive sequences for hybridization but can also cause unwanted hybridization or cross-hybridization. Cross-hybridization occurs when a given target shares significant sequence identity ($> 70\%$ over ~ 1000 nucleotides) with several probes in the hybridization mixture.

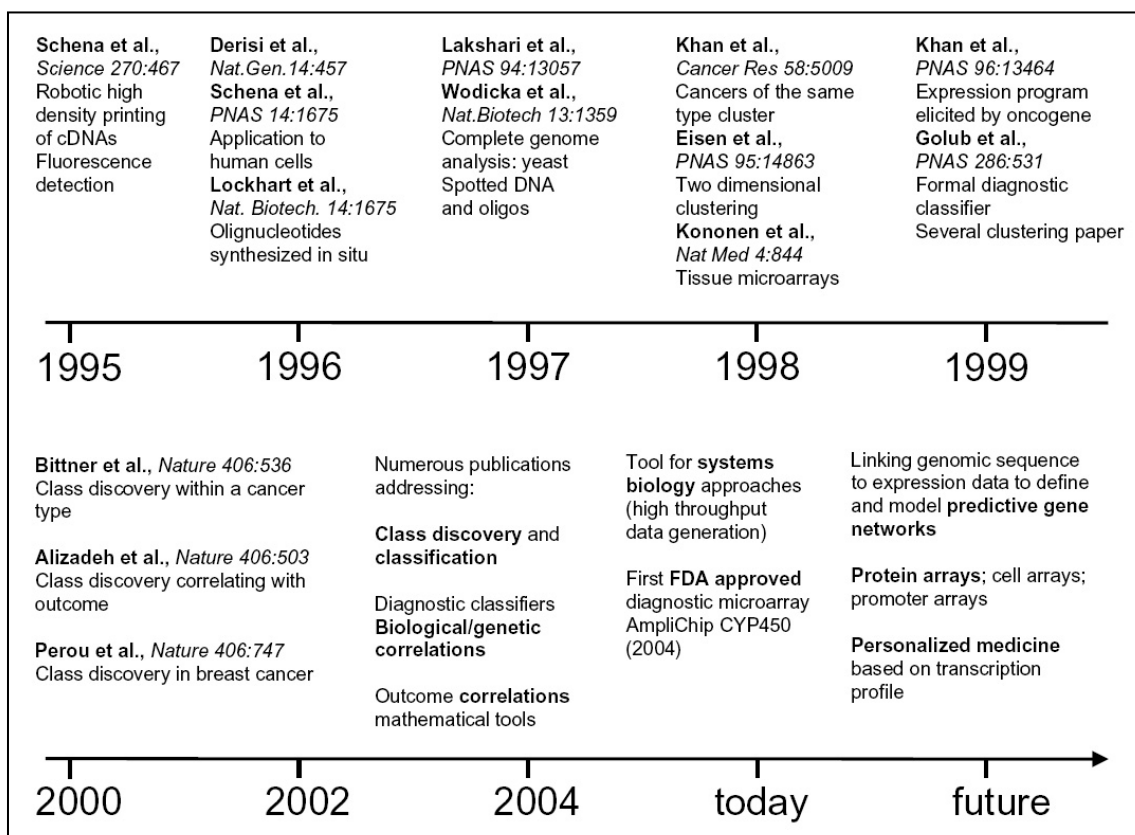


Figure 9: Historical overview of some milestones in microarray development based and extended from an overview of Wei-Hsin Sun.

Oligonucleotide microarrays are another microarray format which is commonly used in many applications including gene expression profiling and genotyping. Oligonucleotides are single stranded 15 - 70 nucleotides long, chemically synthesized molecules. They offer high specificity and good signal strength after hybridization. The main advantage of oligonucleotide targets is increased specificity due to their smaller

size (< 100 nucleotides). More than one quarter of all microarray publications use oligonucleotide microarrays.

In both cases, for complementary DNA as well as oligonucleotide arrays the process of hybridization is used to generate a microarray signal. Both kinds of arrays are nucleic acid microarrays, which cover microarrays containing any type of DNA or RNA as a target. In theory, either cDNA as well as oligonucleotides can be used on microarrays as probes, with the nucleic acid sequence representing the gene whose expression level should be monitored. The long cDNA probes have the advantage of very high sensitivity but sometimes show only low specificity. This can lead to cross hybridization, resulting in positive signals for homologous genes. Using probes comprising 500 to 1000 bases it is sometimes challenging to obtain very specific signals for a target gene. However the cross hybridization problem using cDNA targets should not be overemphasized. For most genes, large targets are still highly specific in microarray hybridization experiments. For example BLAST analysis of interleukin 21 (IL21) reveals in both cases a single hit in the entire human genome whether you searched with a 600 nucleotide cDNA or a 70mer oligonucleotide of IL21 (homology data courtesy of the National Centre for Biotechnology Information, Bethesda, MD.). A similar result is obtained for most human genes.

In contrast to this approach the very common oligonucleotide arrays which are only 25 to 80 bases offer a higher specificity, based on their smaller size. Even high homologous sequences from large gene families or splicing variants can be identified uniquely using oligonucleotides. For genotyping applications with single base mismatch detection oligonucleotides are essential. The smaller ones, like 25mer probes can again result in cross hybridization because they are too small to give a specific signal. The longer oligonucleotide arrays are a good compromise between specificity and signal intensity but they are more difficult to synthesize compared to shorter ones. The company Eppendorf tried with a new approach to combine the advantages of high hybridization levels and therefore strong signal intensities of cDNA arrays with the high specificity of long oligonucleotide arrays in one new type of nucleic acid microarray using so called XmerTM probes as target. These XmerTM probes are 200 - 400 bases long and were specially designed to show almost no homology and the

absence of repetitive elements with other genes to minimize cross hybridization. During the design process specificity is checked by comparing the probe sequence with all sequences in public data bases. Finally, the distance to the 3'-poly-(A) tail is optimized to ensure that also smaller reverse transcription products could be covered with these capture probes. After this *in silico* design, each probe is tested in *in situ* experiments to check again for cross hybridization (deLongueville and Heim, 2004). This new approach for probe design leads to very good performance of this new kind of nucleic acid arrays as indicated in Figure 10.

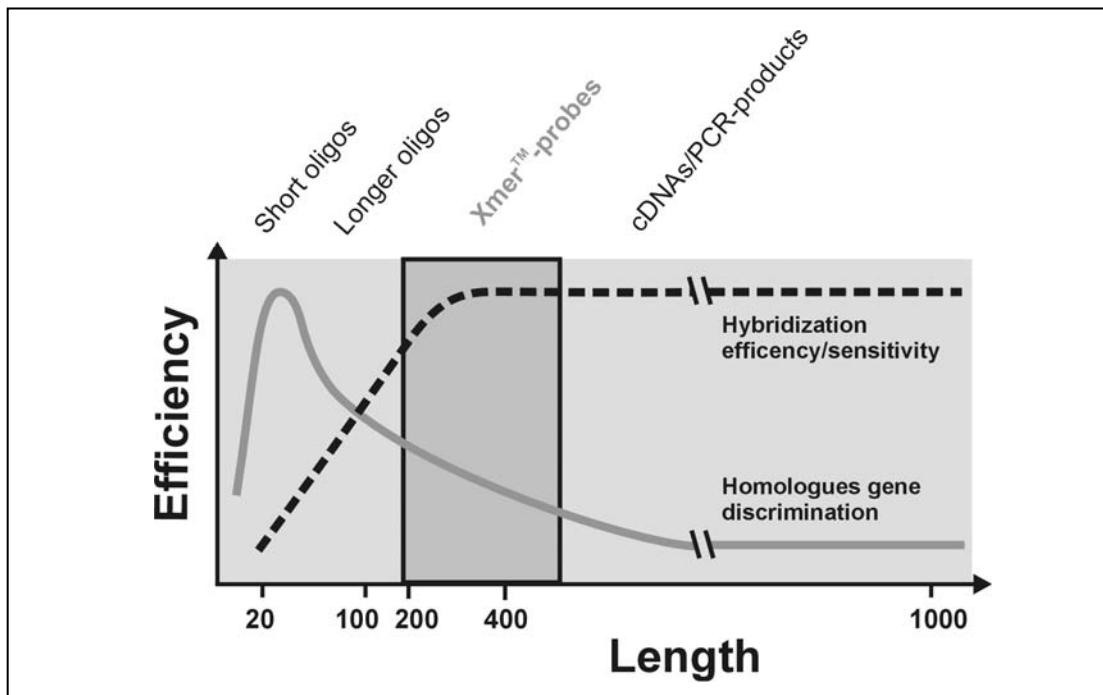


Figure 10: Hybridization efficiency and gene discrimination depending on probe length based on deLongueville and Heim, 2004. The Xmer™ probes (dark grey field) combine the excellent hybridization efficiency of a cDNA fragment (dashed line) with the ability of an oligonucleotide to distinguish homologous genes (solid grey line).

Beside nucleic acid arrays there are several other kinds of microarrays in use right now. Figure 11 shows an overview of microarray formats used and species analysed with this technology. The numbers given here are based on 2,000 microarray citations (Skena, 2003). Tissue and protein microarrays are more or less in the research phase but are found in the literature with an increasing frequency and count for about 10 % of scientific publications (Figure 11 A). This new formats make several classic histological or biochemical assays obsolete. The highly parallel, miniaturized and

automated procedures in microarray analysis give a higher precision, speed and information content which makes these assays superior to the older ones. Tissue microarrays contain sections from tissue specimens like e.g. different kind of tumours and protein microarrays contain pure proteins, cell extracts or antibodies against specific proteins on each microarray location.

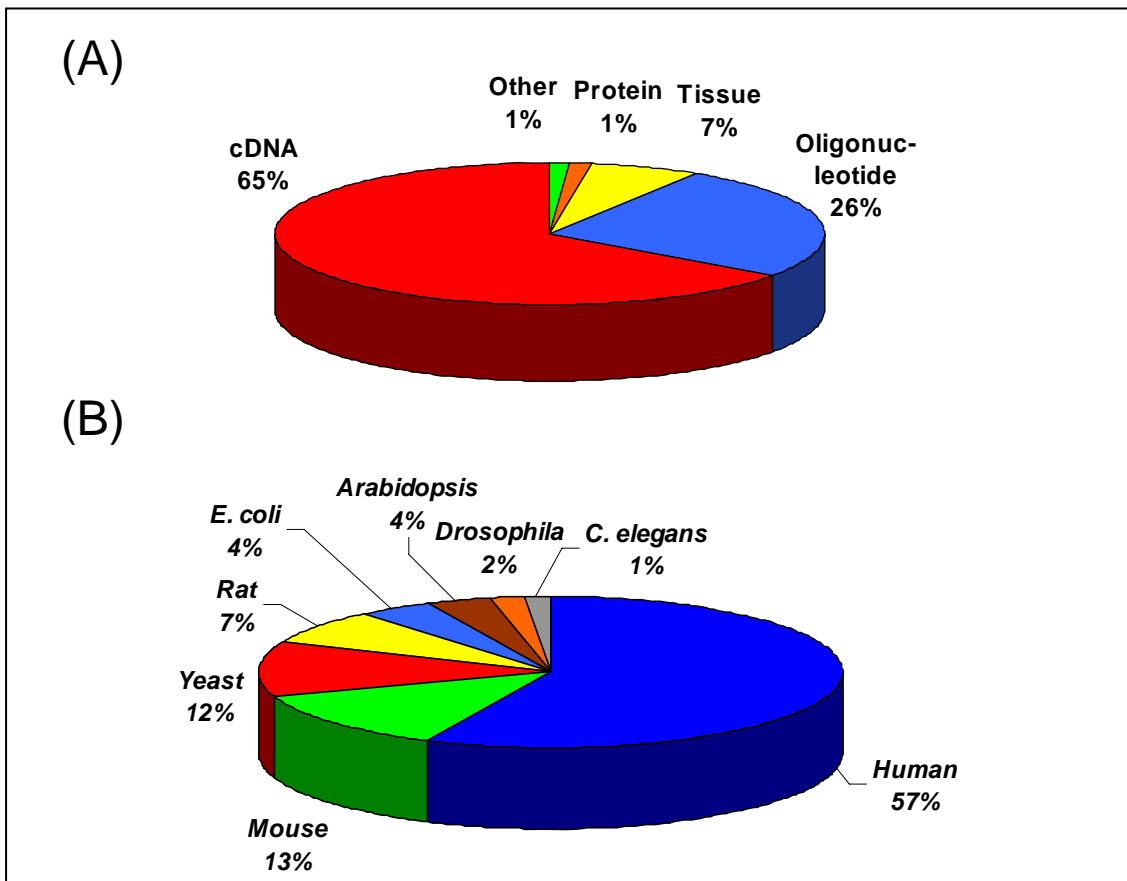


Figure 11: Different kinds of microarray formats. (A) Under the different types of microarrays the cDNA arrays are with 65 % still the most commonly used ones. Protein and tissue arrays play a minor role. (B) Distribution of species analysed with microarrays. Man (57 %) is the most commonly investigated species using DNA microarrays followed by mouse (13 %), yeast (12 %) and rat (7 %).

Next to humans which account for more than half of all publications (57 %), three other organisms - mouse (13 %), yeast (12 %) and rat (7 %) - make up for nearly one third of the microarray citations (Figure 11 B). The common laboratory bacterium *E. coli* which was studied recently by Lemuth, 2006, the model plant *Arabidopsis thaliana*, the fruit fly (*Drosophila melanogaster*) and the worm (*Caenorhabditis elegans*) account for 11 % of the total publications.

These figures underscore the diversity of organisms that are being studied by microarrays. Meanwhile a lot of research is going on in the detection of species in medical and environmental samples (like in the intestine or for waste-water treatment) and also production hosts (like *Pseudomonas* or *Bacillus*). The breakthrough in these fields is expected to come in the next few years.

The amount of publications is still almost exponentially increasing (Figure 12). In addition to the use for the analysis of gene expression (which is with 81.5 % the main application so far), microarrays are also used for other applications like detection of genetic aberrations and polymorphisms.

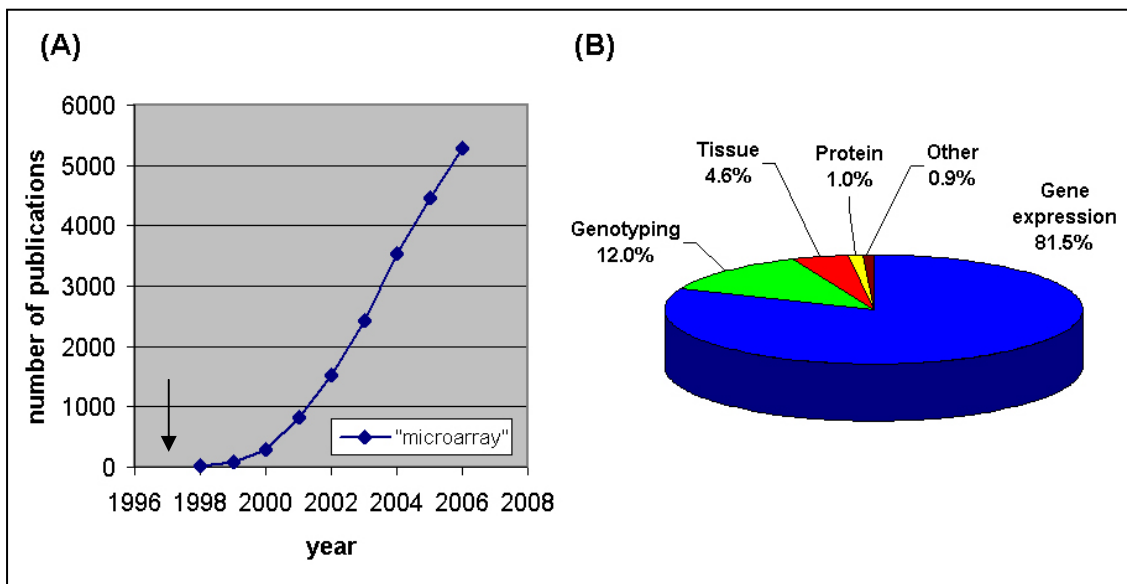


Figure 12: Popularity and usage of Microarrays. (A) Total number of scientific microarray papers published and listed in the literature database "pubmed" as function of calendar year. Beginning with the first publication in Science in 1995 the increase started especially after 1997 when the first commercial microarray robots and scanners were introduced 1997 (arrow). (B) Microarray papers categorized according to research application. The majority of the published microarray studies is still in the field of gene expression (81.5 %) followed by genotyping (12.0 %). The other applications like tissue-, protein- and other kind of arrays (6.5 %) are steadily increasing but still play a minor role.

A quite common application (12 %) is the detection of single nucleotide polymorphisms (SNPs). The only DNA microarray getting an FDA approval so far is a genotyping array from Roche for the detection of SNPs in drug metabolizing enzymes.

Essentially every cell in a multi cellular organism contains an identical copy of genetic information (DNA). However concerning shape, size and function cells differ

dramatically. One of the reasons for these differences is the genetic programming which alters the kind and number of genes expressed in different cells at different time points. By examining gene expression patterns on a genome wide level, microarrays can be used to establish a database for gene expression levels as function of cell, tissue (Figure 13) and time. Such databases are helpful to understand basic mechanisms that control multi cellular development as well as pathological events leading to diseases. The human brain has been the most actively studied human tissue (19 %) and nine other tissues (liver, breast, prostate, lung, colon, kidney, heart, bladder and skin) account for over 80 % of all publications (Figure 13). Building data sets of gene expression profiles for each cell type in the human and other organisms is shedding further light on development, human disease and drug discovery.

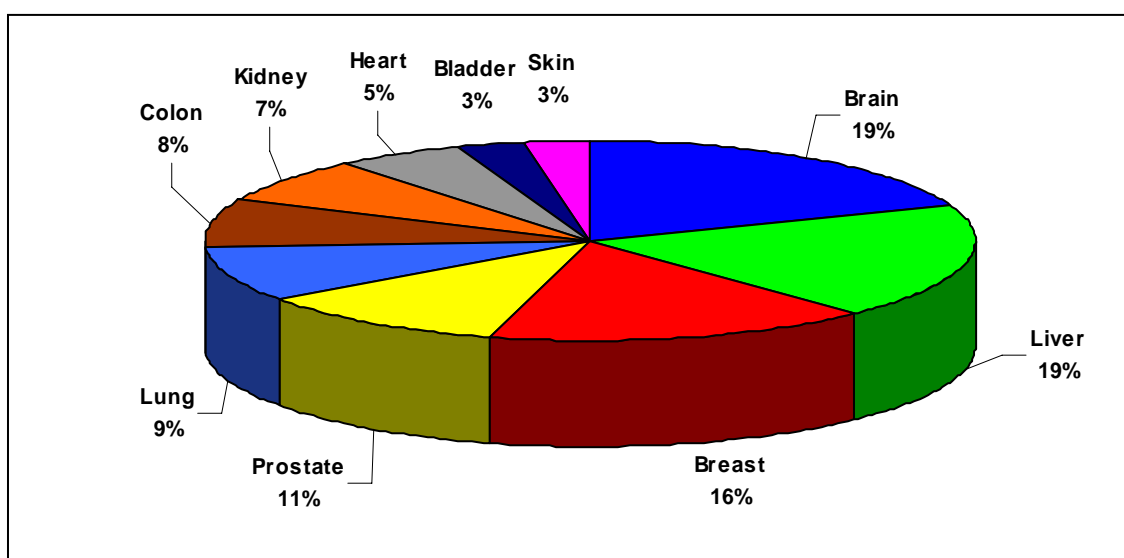


Figure 13: Analysis of microarray publications originating between 1995 and 2001 concerning human tissue. The microarray analysis of brain (19 %), liver (19 %) and breast (16 %) are the most common ones (Data obtained by analysis of the Microarray Electronic library, arrayit.com/e-library).

The beginning and progression of a human disease is based on several factors like genetics, diet and environment as well as the presence of an infecting agent. As one can imagine these factors are quite different and a complex interplay leads to an onset of disease. However all factors have something in common, they can be and most are detected by microarrays. The most important human disease studied with arrays is cancer which accounts for the remarkable number of 83 % of all publications.

Nevertheless diabetes, cardiovascular disease, Alzheimer, stroke, AIDS, cystic fibrosis, Parkinson, autism and anaemia are under intensive investigation by microarrays as well (Figure 14).

By comparing gene expression patterns between ill and healthy people it is possible to elucidate the genetic background of diseases. Knowing the molecular basis of disease enhances the possibility to understand genetic predisposition and will allow individual, save and more effective treatments. All human illness can be studied by microarray analysis and the ultimate goal is to develop treatments or cures for every human disease by 2100.

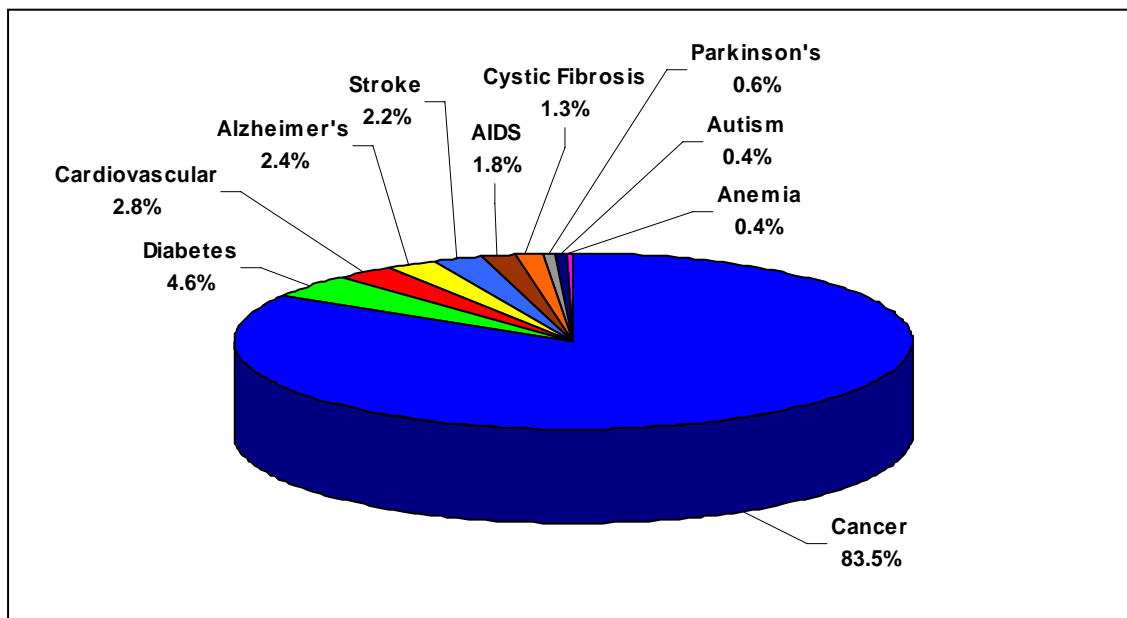


Figure 14: Analysis of microarray publications between 1995 and 2001 categorized according to human diseases. With 83.5 % the investigation of cancer using microarrays is by far the most important, followed by diabetes (4.6 %), cardiovascular illness (2.8 %), Alzheimer's disease (2.4 %) and stroke (2.2 %). AIDS, Cystic Fibrosis (1.3 %), Parkinson (0.6 %), Autism (0.4 %) and Anemia (0.4 %) accounting for together 4.5 % play so far a minor role in investigations using microarrays.

1.4.1 Microarrays in Drug Discovery, Drug Development and Future Applications

Another important field for the use of DNA microarrays is the discovery and development of new drugs. The activity of many drugs is based on their ability to bind to specific cellular targets and alter the expression of specific genes. Monitoring the

expression level of patients undergoing drug treatment would be a helpful tool to support clinical trials, investigate drug dosage and progression of the healing process. Furthermore experiments with cell cultures will allow prediction of molecular targets of potential drugs and will lead to discovery of new drugs that alter the expression of target genes. Drugs changing the expression of non target genes could be identified in the same approach, leading to prevention of side effects. This intelligent approach to drug discovery, where the systems biology project "Hepatosys" will deliver its part, may reduce the cost of drug development and produce safer medicines with fewer side effects. Microarray analysis can also be used for patient genotyping and can classify the population into drug responders and non responders based on the respective genetic background. Therefore this analysis can allow development of personalized medical treatment in the future.

Next to the classical applications like gene expression analysis and biomarker identification in tumours, microarrays also find their way into new fields of application like food analytics (e.g. to detect genetically modified organisms GMOs) or the upcoming field of microarray based species detection in the human intestine in respect to nutrition and health. So finally the microarrays will find their way from the research laboratory into application not only in the clinical environment.

In the following the term microarray is used exclusively for nucleic acid microarrays (oligo- / cDNA arrays) as they were used in this work here, where the expression levels of genes involved in response to a drug administration were monitored and analysed. Before giving a more detailed view on the basic technology of microarrays the following overview shows the structure of an experiment and the workflow to be performed.

Due to cost restrictions for consumables and time constraints the experiment has to be planned very carefully to give maximum information. This planning procedure consists of three parts (Yang and Speed, 2002):

Scientific background: What question is considered? What is the aim of this analysis? How can the experiment answer this question? What is the final output of the experiment?

Practical approach (logistics): Which RNA samples are compared (types, time series, treatment, and control versus reference)? What kind of perturbation is used? What time points have to be selected?

Performance: What kind of sample preparation and labeling (amplification generating aRNA or reverse transcription leading to cDNA, direct or indirect labeling with biotin or fluorescent dyes) should be used? Which (positive and negative) controls are available? How can the results be verified (real-time PCR, Northern Blot)?

The scientific question is usually obvious. More difficult is the question about the experimental design, especially in two colour experiments (Figure 15). The objective of experimental design is to plan the experiment in a way that results are kept as simple and as powerful as possible. The cDNA microarrays mostly have paired hybridization intensity measurements that are taken from two wavelength bands after laser excitation at two wavelengths. These two sources are known as two channels (the Cyanine 3 = Cy3-, green-channel and the Cyanine 5 = Cy5, red-channel). In contrast one colour chips like the ones from Affymetrix are single channel arrays. In case of single channel arrays the two signal intensities of two samples to be compared are obtained from two arrays. In both cases the goal is to assess the relative abundance of RNA in a sample by hybridizing it to a complementary sequence on the array. To avoid technical variance the samples to be compared should be isolated and labeled in parallel on the same day and in two channel experiments hybridized on the same slide. The reason is that there is more variation between slides than within slides (Yang and Speed, 2002). So the most important issue is the question which sample is labeled with which fluorophore and hybridized on the same array. Alternatively it would be possible to compare different samples (points in time) against one reference. However this reference design is more susceptible to a higher variance. To reduce variance it is possible to use a common reference design with time matched controls (for every sample a reference on the same slide is hybridized). To make it more obvious a graphical representation of microarray experimental designs is given in Figure 15.

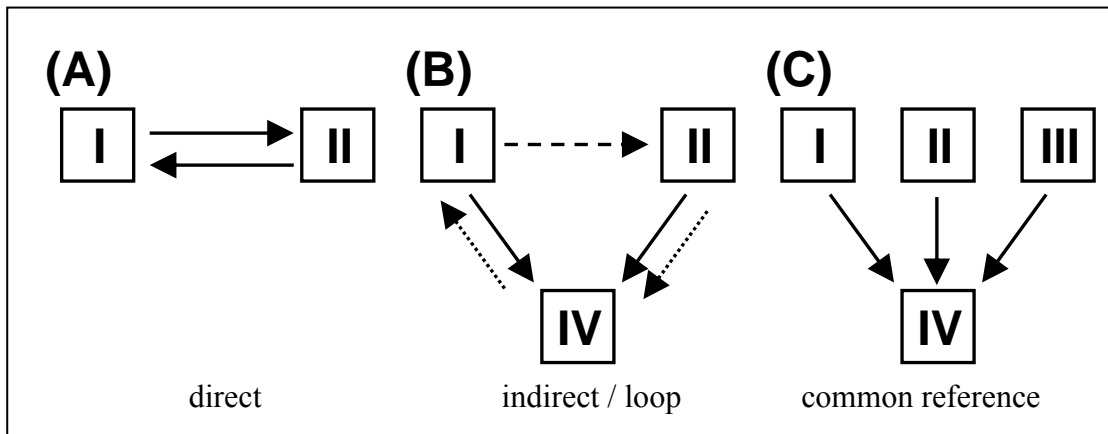


Figure 15: Different experimental designs for comparing gene expression levels in at least two samples. (A) In a single experiment the differential expression of the genes in two samples is compared directly on the same slide. (B) An indirect comparison: sample "I" and "II" are measured separately on two different slides. The log ratio $\log_2("I"/"II")$ is estimated by the difference $\log_2("I"/"IV") - \log_2("II"/"IV")$. The dashed and dotted arrows shown in (B) give the simplest loop design. Three (or more) samples 1,2,3, ... n are hybridized together in pairs in a sequentially way (1,2), (2,3) ... (n-1,n), (n,1). (C) The common reference design is suitable if samples are compared to a reference to identify differentially expressed genes or common regulated genes. "I", "II", "III" = target mRNA, "IV" = reference, arrows correspond to hybridizations between two mRNA samples (based on Yang and Speed, 2002).

The structure of the graph indicates which expression differences in the samples ("I", "II" and "III") can be estimated. Gene expression differences can only be compared if they are linked together by a common (hybridization) experiment (indicated by an arrow). The precision of this determination is given by the number of arrows (experiments) and length of paths connecting each sample to be compared. In the loop design in Figure 15 B for example there are three hybridisations of sample "I", "II" and "IV". Two ways (arrows) connect "I" and "II". Path one joins "I" and "II" directly and path two joins "I" and "II" in two steps via "IV". When the relative abundance of mRNA between target samples "I" and "II" is calculated, the estimate of the log ratio ("I"/"II") from the direct path "I" to "II" (co-hybridization on the same array, dashed arrow) is likely to be more precise than the indirect estimate $\log("I"/"II") = \log("I"/"IV") - \log("IV"/"II")$ from the path of length two that joins "I" and "II" through "IV" (from the two experiments in which "I" is co-hybridized with "IV" and "II" is co-hybridized with "IV" separately, dotted arrows). However the simplest way to compare two samples as indicated in Figure 15 A would be direct comparison. There are two hybridizations (two arrows) comparing "I" with "II" thus two independent estimates of

the $\log("I"/"II")$ ratio for the expressed genes are obtained. If the variance is σ^2 , then the variance of the average of two independent measurements is $\sigma^2/2$. For the indirect design (Figure 15 B solid arrows) with a common reference ("IV") and the hybridizations "I"-IV and "II"-IV the log ratio $\log("I"/"II")$ is the difference of two independent log ratios $\log("I"/"IV")$ and $\log("II"/"IV")$: $\log("I"/"II") = \log("I"/"IV") - \log("II"/"IV")$ as given before. The variance for a single log ratio is again σ^2 . The variance of the difference of two independent log ratios is therefore $2 \sigma^2$.

To summarize, with two hybridizations (solid arrows) using a direct comparison the log ratio of a gene with a variance of $\sigma^2/2$ is obtained, while the variance would be $2 \sigma^2$ using indirect comparison. The fact that the variance differs by a factor of four shows the importance of choosing an appropriate design. Based on these considerations and further general issues discussed previously like the scientific aims and experimental constraints a suitable design has to be developed and applied.

After determining the experimental setup, the number and time points needed to answer a certain question are known. Following these considerations the number of samples, e.g. the amount of necessary cell cultures can be estimated.

The practical part starts with a cell culture which has to be analysed. This culture is often split into two parts, one reference and one perturbed culture. The expression level in both cultures is then compared. Another set up would be to compare the same culture after different time points to investigate changes due to time or analyse the differences between different cells, like different tumour types or cells from different organs or species. In all cases the RNA has to be extracted from the cells. This RNA is directly reverse transcribed into cDNA or amplified using an in vitro transcription protocol (Van Gelder *et al.*, 1990). In both of these steps the nucleic acid has to be labeled and thus can be detected on the array. During hybridization the labeled samples, if expressed in the cells, bind to the complementary sequence which is present on the array resulting in a fluorescent signal. After this hybridization this signal is obtained during the scanning process. The light signals are quantified and background corrected and used for further analysis. The spot intensity measured is indicative for the relative abundance of mRNA in the samples, corresponding to the DNA probe on this array position (Figure 16).

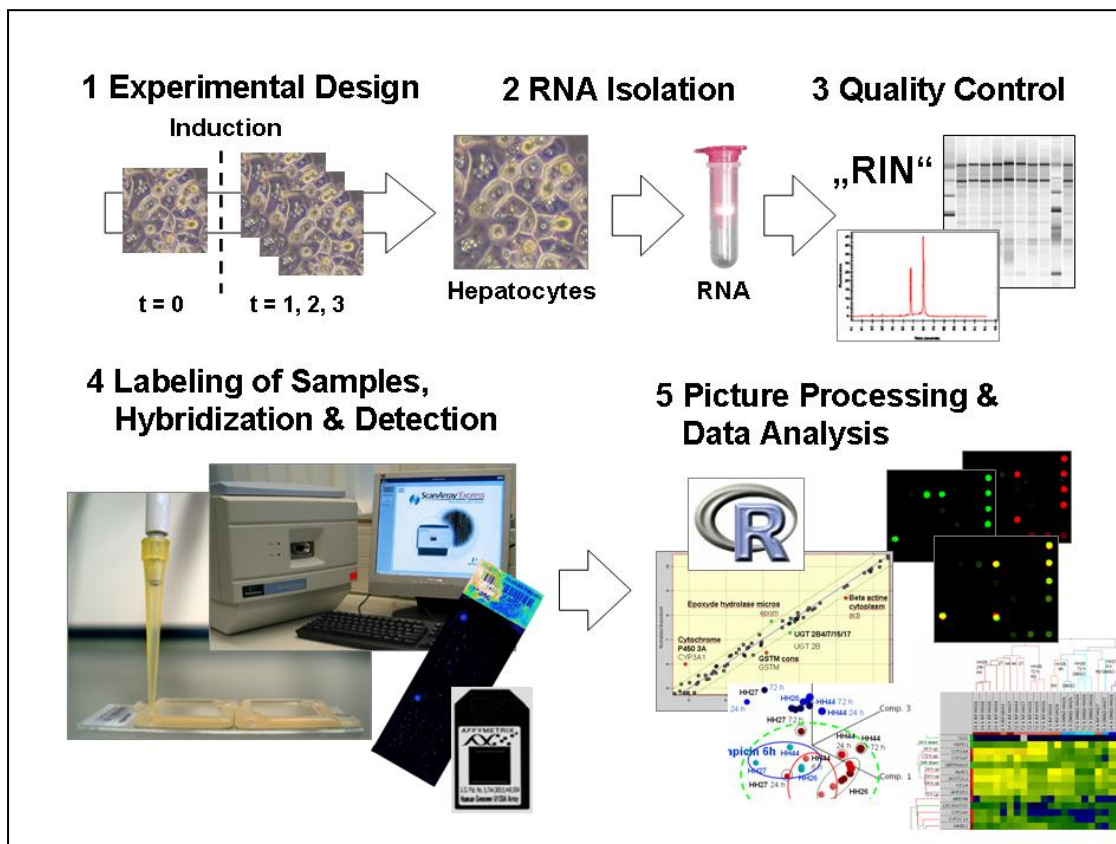


Figure 16: Workflow of a microarray experiment. (1) In the first step the experimental design is planned. According to this design cells are cultivated and perturbed with a stimulus. At e.g. different time points or under different conditions (like induced and not induced) samples are collected and the total RNA is isolated (2). (3) The RNA is checked for its quality and integrity and if the RNA integrity number (RIN) is appropriate (> 7.5) the array experiment is performed. (4) Samples are labeled during amplification (aRNA) or reverse transcription (cDNA) and hybridized on the array. The microarray is scanned and the signal intensity data is quantified and further analysed to obtain expression values (5).

1.4.2 Basics of Microarray Technology

As mentioned previously microarrays can be divided into three major groups according to the type of capture probes used: with short oligos, with long oligos or PCR products (cDNA). All three kinds allow expression profiling experiments which are the major application of microarrays so far. The arrays of the company Eppendorf are a specialized format of cDNA arrays.

For an expression profiling experiment DNA probes representing a gene of the organism to be investigated are immobilized on a surface. This so called "spotting", the

robot assisted application of specific oligo-DNAs on predefined places on e.g. activated glass surfaces is often done by core-facilities or companies. Alternatively probes can be synthesized directly on the microarray using light inducible surfaces and mask protection technology (Lipshutz *et al.*, 1999; McGall and Christians, 2002). Using this manufacturing technology more than a million different probes can be synthesized on an array roughly the size of a thumbnail in a few hours.

It is important that the probes used are specific to avoid cross hybridization. Therefore an oligonucleotide of 50 bases should have less than 75% sequence homology to all other genes (Kane *et al.*, 2000). To provide optimal hybridization conditions the guanine (G) cytosine (C) content and melting temperature should be optimized and probes should have minimized secondary structures and show no self-self dimerization (Southern *et al.*, 1999). When the arrays with the immobilized probes are hybridized with a sample, nucleic acids from the sample will bind to complementary probes on the array. In an expression analysis experiment usually two states (or more) are compared e.g. treated or not treated, before and after induction, at two different time points and so on, so there are two, test- and reference samples hybridized. This can be done on one hand using one colour arrays where the two samples are hybridized on two separate areas of one slide (like the Eppendorf DualChips®) or on two different slides (Affymetrix). On the other hand hybridization on two colour arrays (Eppendorf, Agilent) can be done, allowing both samples to be analysed in the same area on one slide. In this case samples can be discriminated by a different label. The comparison of gene expression in two samples leads to relative transcription levels (Yang and Speed, 2002). Main parts of this work were done using Eppendorf DualChips® and Affymetrix full genome chips. Therefore these two platforms will be introduced in detail in the following part. As mentioned previously the DualChips® contain cDNA like oligo probes (Xmer™) on their surface allowing a good discrimination and yielding in a high signal to noise ratio (de Longueville *et al.*, 2003). The probes for the 151 genes represented on this array are spotted with a robot on two distinct areas of one glass slide in triplicates. A complementary probe for every gene is represented six times on one slide, three times for each of the two (reference and test) samples (see Figure 17).

This special format allows one colour as well as two colour hybridisation experiments. In the case of a one colour experiment both samples are equally labeled and hybridized

on two separate areas containing similar probes on one slide. For two colour experiments where the two samples are labeled with different dyes both samples could be analysed on the same probe area because the complementary nucleic acids in the mixture will compete for the hybridization on the probes. The sample with the higher amount of nucleic acid will hybridize more often yielding in a proportionally stronger signal. As the two samples have to be labeled differently it is advisable to use the other hybridization area for a so called dye swap experiment. During such an experiment the label is reversed, so if in the first case the test sample was labeled with dye 1 and the reference with dye 2 then in the swapped experiment the test sample is labeled with dye 2 and the reference with dye 1. The dye swap allows correcting signals for unequal labeling and hybridisation.

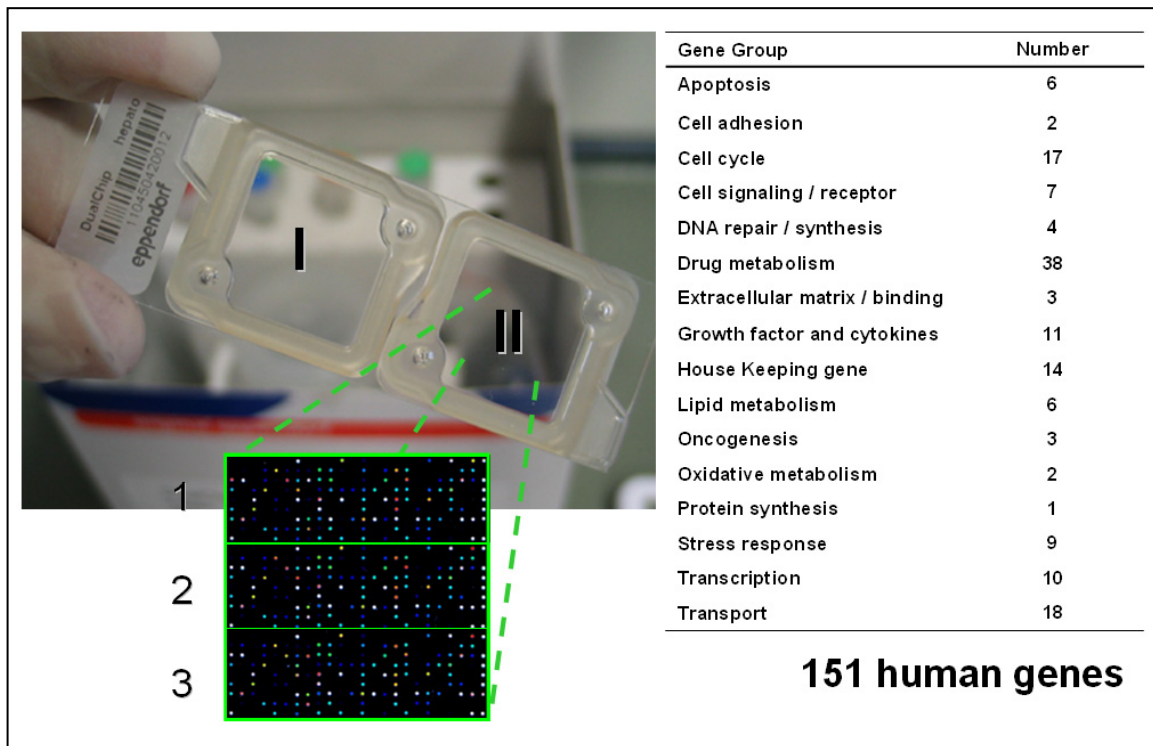


Figure 17: Design of Eppendorf DualChip® Microarrays. The probes for 151 human genes of different biological functions are spotted two times on separate hybridization areas (I, II) in triplicates (1, 2, and 3) on a standard glass slide. The spotted areas are already sealed with a hybridization chamber.

Labeling can be done using different approaches. One approach uses fluorescent molecules like Cyanin-3 (Cy3) and Cyanin-5 (Cy5) which are incorporated in the sample during reverse transcription PCR converting the unstable mRNA target into a complementary cDNA sample. For this reaction modified nucleotides carrying a

fluorescent dye are used, resulting in a fluorescently labeled cDNA which can be hybridized on the array. This procedure is the "direct labeling" method where the fluorescent dye is directly incorporated into the target molecule. As the fluorescent dye is big and therefore the labeling rate sometimes can be quite low, it has to be optimized by trying different ratios of labeled and unlabeled nucleotides and temperatures. In the case of direct labeling with two different fluorescent dyes and mixing both samples a competitive hybridization on the microarray occurs. The two samples of target compete for binding on the same collection of probe. If cDNA in one sample is predominant the final signal has the colour of this sample (false colour: red Cy5, or green Cy3) or if the amount of cDNA in both samples is approximately equal then it would result in a yellow colour (mixture of both colours). A graphical overview of a microarray experiment using direct labeling is given in Figure 18.

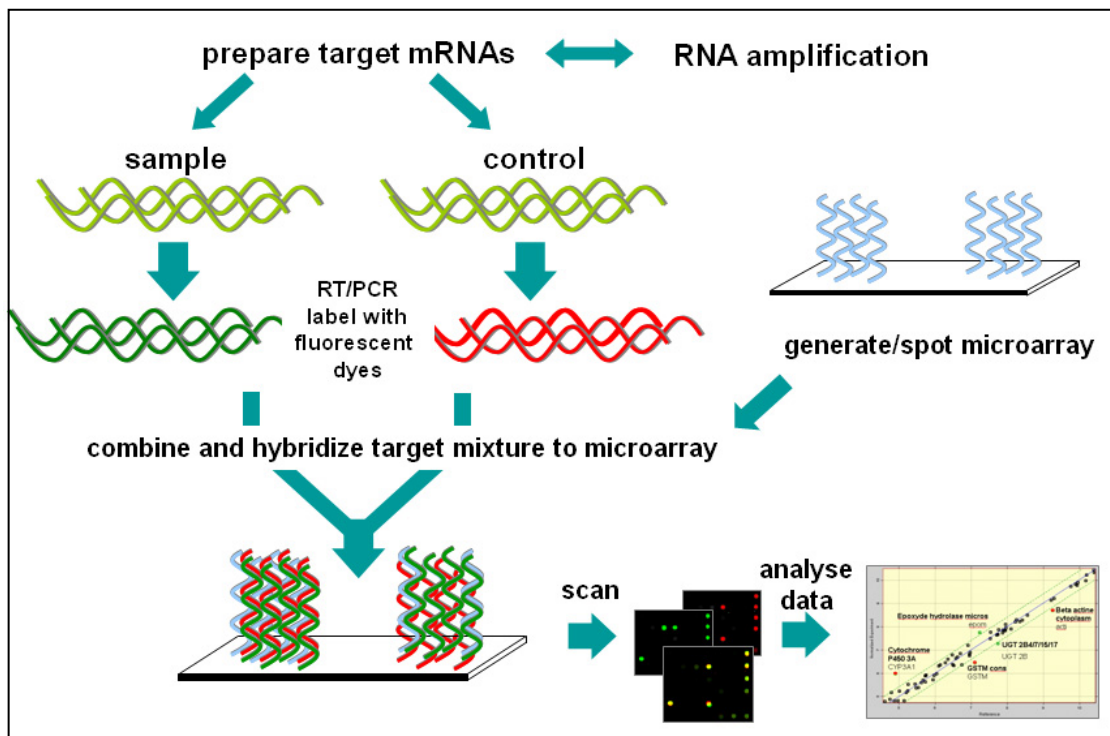


Figure 18: Experimental procedure of a two colour direct labeling microarray experiment. A control RNA and a sample RNA are reversely transcribed and labeled differently. Both are mixed together and hybridized on the same array. After scanning a false colour image is obtained with either one colour (red), the other colour (green) or a mixture of both (yellow) depending on the expression level of a gene in the different samples. These images are quantified and the data is then analysed further.

Another alternative is the indirect labeling method (Figure 19). Using this strategy e.g. biotin-labeled nucleotides which are less bulky are used to mark the cDNA. As biotin

itself gives no fluorescent signal a second detection step is necessary. During this step the hybridized sample on the array is detected by using a fluorescence labeled anti-biotin-antibody which is capable to bind the cDNA due to the incorporated biotin-label. This detection step increases sensitivity as several antibodies can bind to biotin molecules in one oligonucleotide. Detection is done using a fluorescence scanner and both samples are quantified yielding absolute expression data of the sample hybridized on this array. The comparison and calculation of expression ratios is then done in silico using mathematical algorithms.

In contrast to direct two colour labeling experiments in a one colour labeling experiment there is no competitive hybridization on the array and therefore the values obtained are not inherently comparative. The comparison of two samples hybridized on two arrays is done after quantification and normalization mathematically. The final result is the relative abundance of two sets of mRNA.

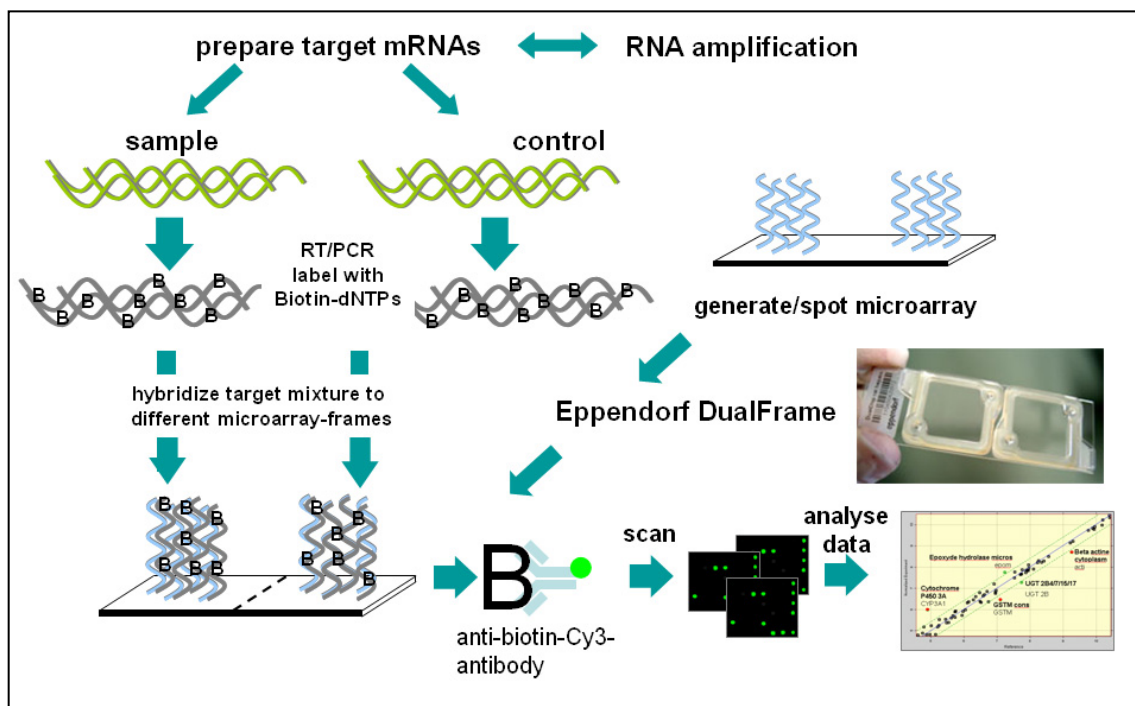


Figure 19: Experimental procedure of a one colour, indirect labeling microarray experiment using DualChips® from the company Eppendorf. Here both samples are labeled similarly. Therefore they can not be mixed and are hybridized on two separate arrays. The obtained fluorescence values can be assigned to each sample based on the array they were hybridized on.

The same labeling method (indirect biotin labeling) is used for the Affymetrix GeneChips® used in this work. The main differences of the unique Affymetrix Chip technology is (1) the idea of using perfect match (PM) and mismatch probes (MM) to obtain the background corrected signal, (2) the size and (3) the fabrication of these arrays. The basic experimental setup for the array experiment shown in Figure 20 is the same.

As stated previously the Affymetrix GeneChips® are not spotted microarrays, as it is common when using standard glass slides. The special format is build using photolithographic DNA synthesis. Using methods from the semiconductor industry, GeneChip® array manufacturing starts with a silane modified quartz wafer (Fodor *et al.*, 1991; McGall and Christians, 2002).

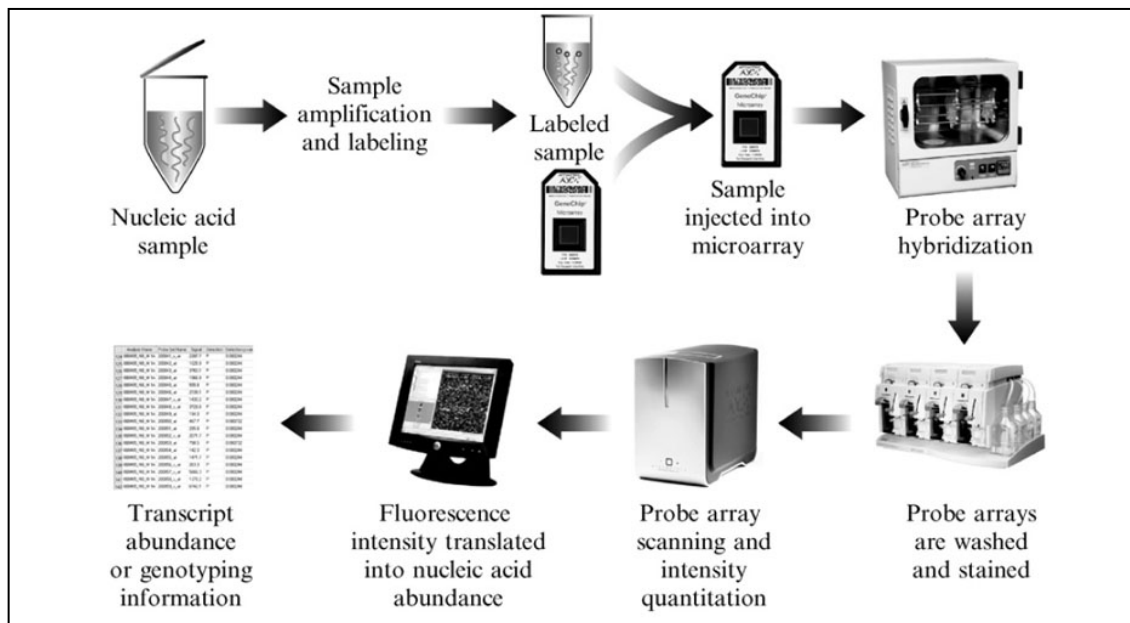


Figure 20: Flow chart of an Affymetrix GeneChip® microarray experiment. The special format of the GeneChips® requires special hardware for hybridization, washing and scanning. Otherwise the principle is the same like during a direct labeling one colour microarray experiment on standard glass slides. After preparation of the sample it is amplified and labeled. The labeled sample is hybridized on the array in a hybridization oven followed by the washing procedure. Finally the washed arrays are scanned and the fluorescent intensity for every feature on the array is obtained. Data output contains an intensity measurement for each transcript or genotyping (SNP) information (graph: Dennise *et al.*, 2006).

This modification provides stable surface layers of hydroxyalkylgroups where photo labile protection groups can be attached. These surfaces can be activated by light (Figure 21 A). A photolithographic mask which exposes certain areas of the surface to UV-light allows the exact activation of specific spots to react (Pease *et al.*, 1994; Gao *et*

al., 2001). By activating the spots in a sequence based order a gene dependent probe is generated. The mask has to be exact aligned to the quartz wafer to ensure that oligo synthesis is only activated at precise locations on the wafer. Ultraviolet light shines through the mask deprotecting terminal hydroxylgroups on the linker molecules in exposed areas. By this activation hydroxyl groups are prepared for nucleotide coupling.

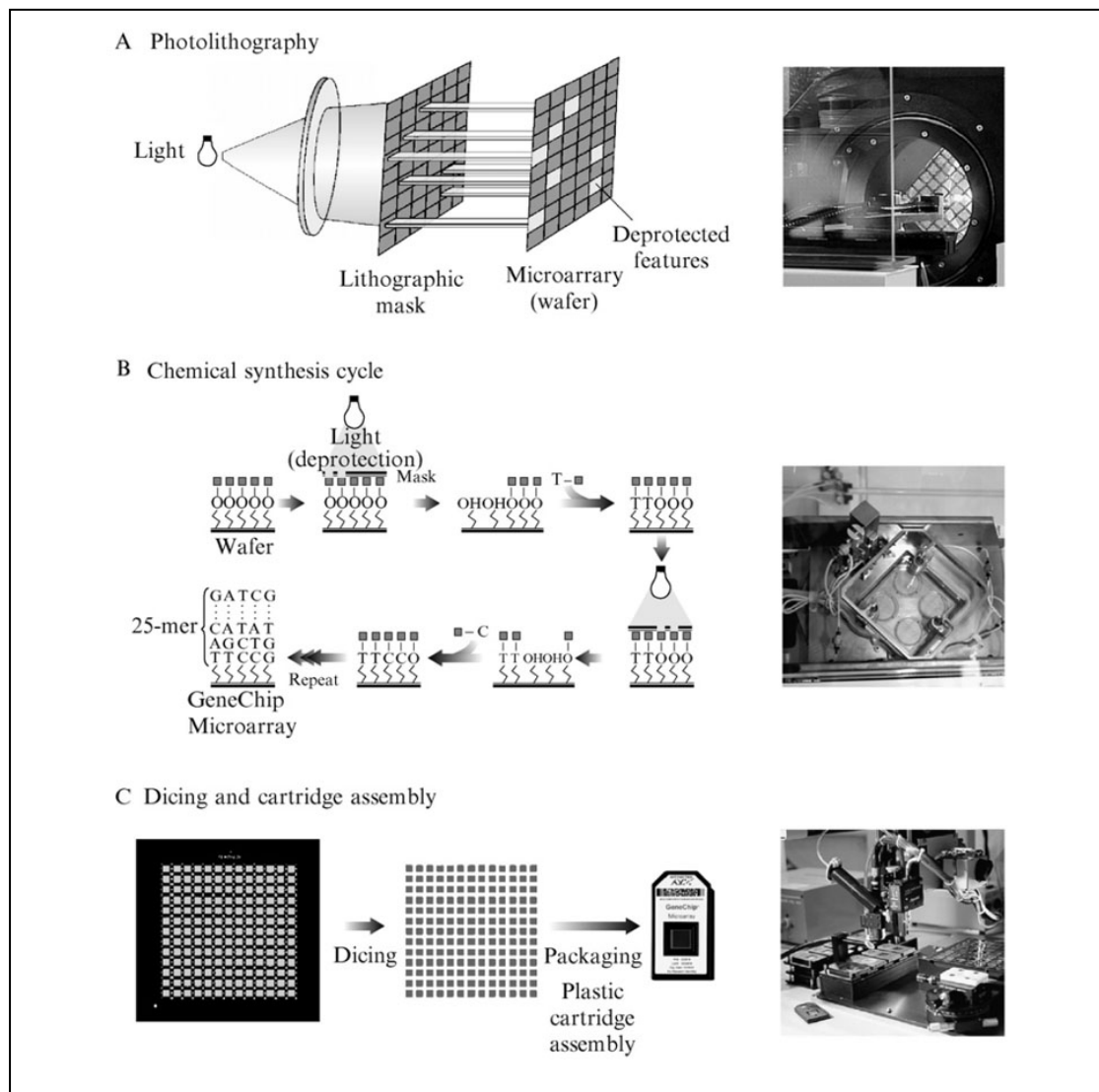


Figure 21: Manufacturing of Affymetrix GeneChips®. On the left side schematic representations of the different steps and on the right side a picture of the hardware used is given. (A) Principle of mask aided photolithographic synthesis of oligos on a surface. (B) Schematic overview of the synthesis cycles consisting of deprotection and coupling with the addition of different nucleotides. (C) Final stages of manufacturing consist of dicing of the wafer into individual GeneChips® which are inserted into plastic cartridges. Graph from Dennise *et al.*, 2006

In contrast, areas that are not exposed remain protected and inactive. A solution containing a deoxynucleoside phosphoramidite monomer with a light sensitive protecting group is passed over the surface of the wafer. At the mask dependent light activated positions the nucleoside attaches to the hydroxyl groups initiating the oligo synthesis. By repeating these two steps of deprotection and coupling the synthesis proceeds using a unique mask for each round of synthesis. The coupling steps alternate by adding all four kinds of modified nucleotides (A-, C-, G- and T-nucleotide). This allows cycle dependent (activation state) incorporation of every base at any position. The deprotection and coupling cycles are repeated until the complete 25mer probe sequences are synthesized (Figure 21 B).

The mask technology is the main reason why especially custom-made arrays of this kind of chips are very expensive. For every sequence position a unique mask has to be designed. The masks must be very precise and are a major factor in the cost of producing each wafer. Once the masks are made, they can be used over and over again. Therefore the price per array declines for large quantities. Specially developed algorithms help to reduce the number of masks needed. The more than 100 cycles required to synthesize all possible sequences of an array of 25mer oligos can be reduced significantly. A typical 1.28 cm² array contains more than 1.4 million different probe locations, assuming that each feature is synthesized 11 μm from each other. In each feature millions of identical DNA molecules complementary to the desired gene sequence are present. A further increase in numbers of features and a reduction of distance is to be expected in the future. The manufacturing process ends with quality control, dicing of the wafer into several arrays and packing into a plastic cartridge. A graphical overview of manufacturing is given in Figure 21 C.

After describing the manufacturing process a brief description will be given about the array design which is also special for the Affymetrix GeneChip® format. In a classical glass slide microarray experiment only complementary probes for the gene of interest are used. The Affymetrix strategy is to use two different kinds of probes for one gene to improve data quality: the first kind of probe has the exact complementary sequence (perfect match probe -PM) while the second has a single mismatch to the target gene, situated in the middle of the oligonucleotide (mismatch probe - MM).

For gene expression arrays (in vitro transcription assay IVT) the probe selection region lies in the first 600 bases proximal to the polyadenylation site (3'end). The selection of the probe sequence is based on specificity to minimize cross hybridization and maximize linearity, which means that the hybridization intensities respond in a linear way to the relative abundance of the target (Mei *et al.*, 2003). Typically each gene is represented by eleven probe pairs per 600 bp probe selection region, consisting of eleven PM and eleven MM probes, which gives a high crediblens of hybridization signals (Figure 22).

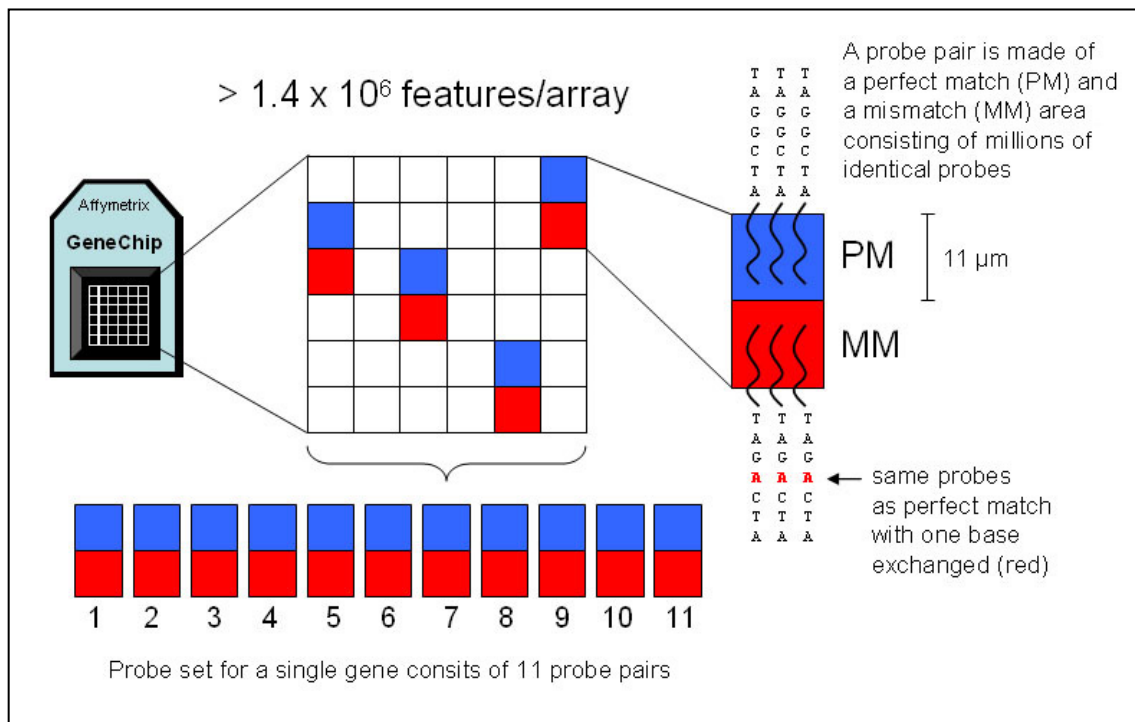


Figure 22: Array design of Affymetrix GeneChip®. In the plastic container of an Affymetrix GeneChip® there is a 1.28 cm² quartz support containing more than a million features. Each feature is composed of millions of complementary oligo nucleotide sequences. Each gene is represented by a probe set which consists of 11 probe pairs distributed randomly on the array. A probe pair is made of a perfect match (PM - blue) and a mismatch (MM - red) probe which is identical to the PM sequence, except for a nucleotide exchange in the middle of the sequence.

Following hybridization, staining with a fluorescent streptavidin-phycoerythrin conjugate (SAPE) to label the biotinylated samples with a fluorophore and washing which is done in a computer controlled fluidics station the arrays are scanned. The scanner is a wide-field, epifluorescent, confocal microscope that uses a laser to excite the bound fluorophores on the chip. The pixel resolution is 0.7 μ m and the latest scanner is able to scan features with 5 μ m spacing with 65,000 different fluorescent

intensities. A photomultiplier tube collects and converts fluorescence values into an electronic signal, which is then converted into the corresponding values.

Using the MM probes for correcting the signal of the PM probes a corrected mean value for each feature can be calculated. For the combination of the eleven different signals there are several algorithms available, like microarray suite 5 (MAS5), robust multichip analysis (RMA) or GC-content dependent robust multichip analysis (GC-RMA). All these algorithms summarize multiple probe intensities into an aggregate signal estimate that is correlated to the relative abundance of the transcript in the experimental sample.

In this work two quite different array platforms were used to monitor gene expression. On one hand the X-merTM (cDNA) based microarray DualChip[®] from the company Eppendorf which allows detection of a subset of liver specific genes (151) with one- and two colour direct and indirect labeling was used. On the other hand the expression of all human genes (30,000) was monitored using a one colour "indirect labeling" approach with GeneChips from Affymetrix.

It had to be clarified whether the results of these quite different platforms can be compared. Next to satisfying experience obtained in this work, which will be discussed in detail in the results part, there are several papers published addressing this question.

The groups discussed comparing different platforms like cDNA microarrays with long oligonucleotide microarrays or Affymetrix GeneChips - with short oligonucleotide microarrays, as well as matching the results from different laboratories performing the analysis (Bammler *et al.*, 2005; Irizarry *et al.*, 2005). A good overview is given in the "microarray quality control (MAQC) project" published in nature biotechnology (Shi *et al.*, 2006). It can be concluded that the impact of different laboratories and scientists performing the experiment is more important than the effect of using different platforms. Nevertheless there are differences found which can be explained by different probe sequences on different platforms for the same target. Therefore studies with different platforms having different capture probes for the same target might show different results. The ability to reproduce results for similar probes on different platforms is quite high. Furthermore the quality between different laboratories can be improved by using standardized protocols (SOPs) for all steps from the experimental procedures to quantification, data correction and analysis.

To obtain a better standardization and control of microarray experiments and to make microarray experiments more comparable to each other, an international organisation (Europe, America and Japan) of biologists, computer scientists and data analysts was founded. The microarray gene expression data (MGED) society focuses on establishing standards for microarray data annotation as well as exchanging and facilitating the creation of microarray databases and related software to implement these standards. The objective is promotion and sharing of high quality, well annotated data within the life sciences community. A long-term goal is to extend the mission to other functional genomics and proteomics high throughput technologies (www.mged.org) (Causton and Game, 2003; Ball *et al.*, 2004; Ball and Brazma, 2006).

The three main concepts of this initiative are:

MIAME (Minimum Information about a Microarray Experiment) is a document giving the necessary information that should be reported about a microarray experiment to enable interpretation and reproduction.

MAGE (Microarray Gene Expression) consists of three parts: MAGE Object model (MAGE-OM) an XML-based document exchange format (MAGE-ML) which is derived from the object model and finally the support of software tools (MAGEstk)

MO (MGED Ontology) defines sets of common terms and annotation rules for microarray experiments, enabling annotation, data analysis and data exchange including all related information

The presented concepts will help to handle the huge amount of biomedical data generated by microarray experiments. Especially for systems biology approaches, where several groups are involved, well defined standards are imperative. The rules for microarray experiments have to be applied also to other high throughput technologies like proteomics and metabolomics to ensure comparability and establish certain standards for public databases. Standards for common parts of experiments (sample origin, manipulation) have to be described, especially when using different technologies. There is a lot of progress since the initial meeting of MGED society in 1999 followed by several international meetings in Heidelberg 2000, Stanford 2001,

Boston 2002, Tokyo 2002 and finally 2006 in Seattle. However, bringing all this information together in a well structured, comprehensive and reproducible way will be a challenge for scientists in the next five to ten years.

To assure good quality data from microarray experiments requires not only good laboratory practice, but also standardized and suitable data processing and analysis. So finally after performing the lab work the time consuming and challenging task of interpretation of "spot intensities" begins.

1.4.3 Data Quantification and Interpretation of Microarray Experiments

The speed and ease of quantification is one of the big advantages of microarray technology. Biological information for thousands of genes can be obtained in a few hours. The simplicity of data acquisition should not deceive the sophisticated theory behind quantification, ratio calculations, log transformation, normalization and other aspects of microarray data quantification.

Microarray scanners deliver images (tiff) of the two dimensional distribution of fluorescent intensity on the microarray surface. Each gene is represented by a spot which consists of several pixels. The process of obtaining numerical values for every spot on the image file is called quantification. The extracted numerical values provide information on the amount of mRNA (or other biological molecules depending on the array) in the hybridized sample. There are two ways to quantify: the absolute quantification yields in numerical values for every spot (gene) based on the signal intensity at each location (number of counts). In contrast relative quantification is the measure of the ratio of two samples on a microarray and calculating the quotient of one channel divided by the other channel (in two colour labeling experiments). Relative quantification results in a quotient of expression between two samples or a so called "fold change". A gene for example is 2-fold higher expressed in sample "A" compared to sample "B", if the signal in "A" is twice as high as in signal "B".

The quantification can be accomplished manually as well as automatically. Nowadays, with the increasing number of spots $> 10,000$, quantification is mainly done automatically. During this procedure a grid, allowing quantification, is placed on the tiff-image specifying the number of rows and columns and assigning each position to a

specific gene. Following the right positioning of the grid to the array, the computer is enabled to quantify each element in this quantization template automatically. Each element consists of the intensity of the spot signal as well as the local background. The process by which microarray signals are delimited from the background is called signal segmentation. The two main forms are spatial and intensity segmentation. Spatial segmentation relies on careful positioning of the grid (quantification boundaries), while intensity segmentation uses pixel intensity values to separate signal from background. Two methods for intensity segmentation are watershed method (Beucher and Meyer, 1993) and seeded region growing (Adams and Bischof, 1994). For both methods a starting point is needed, therefore the spatial segmentation should be done carefully. Starting from this seed pixel, pixels are progressively adjoined and added to the spot until adjacent spots show distinctly lower fluorescence indicating the background. Sophisticated commercial quantization packages like Imagene (BioDiscovery) or ScanArrayExpress (Perkin Elmer - used in this work) incorporate both segmentations in the automatic quantification procedure to obtain highly precise microarray data. Subsequently background correction can be applied. By subtracting the fluorescence signal from the background or from negative control spots, the net fluorescence of the signal corresponding to the amount of RNA in the sample, can be determined.

The data outputs are numerical values for the signal intensities of each spot, the background intensity around the spot, as well as for the background corrected intensity. This data can be stored as tab-, or comma-delimited txt- or CSV-files.

For two colour experiments it is obvious that the outcome is a ratio of the two channels compared on one array. In one colour experiments the signal intensities of two arrays are often divided to obtain a quotient for each gene as for two colour experiments. The so called "ratio calculation" is widely used in microarray analysis, especially for gene expression to obtain a "fold change" as described previously. The absolute signal of two arrays gives a measure of the amount of gene product (mRNA, protein) present in the samples. The extent of the change in concentration of a given gene in two samples provides a measure of gene activation or repression.

Before calculating ratios microarray data is often log-transformed. Data from microarray scanners are expressed in units of fluorescence with 16-bit values ranging from 1 to 65,536. A logarithm (\log_{10}) is defined as the power to which 10 must be

raised to produce a given microarray value. Using a logarithmic scale, 16-bit microarray data are expressed in values ranging from 0 (log 1) to 4.8 (log 65,536). Quantified spots yielding 1,000 and 10,000 fluorescent units would have log transformed values of 3.0 and 4.0. Microarray data transformed into a log scale (0 - 4.8) has a much more uniform distribution than raw signal intensities (1 - 65,536). Log scale data also facilitate the analysis of ratios. Dynamic changes in gene expression are represented as positive (up-regulation) or negative (down-regulation) numbers. Two genes that are induced or repressed by 20-fold and 127-fold would have logarithm values of 1.3 and -2.1. A log scale of -5.0 to +5.0 allows the representation of genes that are repressed by 100,000-fold and activated by 100,000-fold, covering a dynamic range of ten orders of magnitude.

Transformed microarray data are used widely in scatter plots, clustering and other forms of data analysis and visualization (shown in the results part).

The log transformed ratios from two samples are also called M -values. They are calculated by $M = \log_2 X/Y$ (The letter M is a mnemonic for *minus* as $M = \log X - \log Y$ and while A is a mnemonic for *add* as $\text{Log intensity for a spot } A = 1/2(\log X + \log Y)$). Using a base-2 logarithm for M and A is convenient as M is given in units of 2-fold change and A is in units of 2-fold increase in brightness. On this scale, $M = 0$ represents equal expression, $M = 1$ represents a 2-fold change between the RNA samples, $M = 2$ represents a 4-fold change and so on (Smyth *et al.*, 2003).

Due to small changes in the experimental procedure and image acquisition, like different enzyme charges for reverse transcription / amplification, unequal spotting or labeling, variation in laser power, detector sensitivity and others there are minor imbalances between different arrays. To equalize data from different channels or different chips they become "normalized". Normalization can be accomplished using a variety of different criteria, including global intensities, housekeeping genes or internal standards. Global intensity matching works extremely well for samples that share many common signals, like e.g. full genome gene expression of a reference compared to a treated sample. Most of the approximately 30,000 genes would not change there expression pattern, allowing a good normalization. The bias introduced on the array is on a log scale expected to be constantly distributed over the whole array. Therefore the log-ratios can be corrected by subtracting a constant c : $M_{\text{normalized}} = M - c$. For the global

constant c the mean or median of all not differentially expressed genes is chosen. This kind of normalization was applied in the work presented here for the full genome GeneChips from Affymetrix. However for the DualChips® from Eppendorf another approach had to be applied as most of the 151 genes showed differential expression. In this case normalization was done using housekeeping genes, which are known for their more or less equal expression, allowing normalization. Due to strong perturbation or disease also these housekeeping genes might be differentially expressed. Therefore a second normalization step was used, based on an internal standard. An mRNA dilution series of known quantities was spiked into the gene expression labeling procedure. The intensities of the control spots allow the determination of a standard curve based on signals from mRNAs of known quantities. Comparing standard curves and adjusting signals on two different arrays allows also normalization of the arrays.

Finally a quality assessment was performed by checking whether the intensities on the array span the whole dynamic range avoiding saturation. Positive and negative control spots have to be checked for their presence and absence. The background has to be below the signal intensity and the array has to be investigated for spatial effects. Replicated spots were revised for their uniformity and outliers were rejected.

1.4.4 Identifying Differentially Expressed Genes

After quantification and correction of the signals a statistical test was performed to identify differentially expressed genes. The results of a microarray experiment for gene expression can be expressed as a comparison of the expression level of certain genes in two cell samples. For gene i on array j it can be written in general

$$M_{ij} = \log_2 \frac{(\text{expression level in sample 1})_{ij}}{(\text{expression level in sample 2})_{ij}}. \quad (1)$$

The numerator and denominator in (1) are the values from the two different colour channels in a two colour experiment or from two arrays in a one colour experiment. For N genes on each array and n replicates (arrays), the data of an experiment will consist of

$(M_{ij}), i = 1, \dots, N$ and $j = 1, \dots, n$. To determine whether a gene is differentially expressed, it has to be investigated if genes have non zero M -values (Dudoit *et al.*, 2000).

In the beginning of microarray analysis it was just checked whether a certain gene was expressed above or below the mean M -value (average M_i for each gene $i = 1, \dots, N$). M does not depend on the standard error of the genes, so highly variable genes are treated in the same way as stable ones.

The approach using these traditional statistics was quite poor and often not reproducible, because the variability of expression level for every gene was not taken into account. A large mean could e.g. be caused by one outlier. An issue often occurring in microarray experiments (Lönstedt and Speed, 2002). The variability of M -values among genes within replicates is not constant. Genes having a high variance have a high likelihood to get a high M -value, although they are not differentially expressed.

As it is critical to determine whether an observed ratio is due to random or based on different measured quantities, the variation of a measured expression ratio has to be considered. This is done mathematically using a test statistic where the numerator contains the size of the effect (M , ratio of gene intensities). The denominator represents an estimate of the variance, e.g. the standard deviation of the ratio. One of the basic and quite common statistical tests is the Student's t-test. It is the easiest and most straightforward method of calculating whether there is significant difference in expression levels between two samples. If mRNAs from two biological samples "1" and "2" are hybridized onto a small number of arrays (N two colour arrays or $2N$ one colour arrays) the average logged expression ratio is based on M and its sample standard deviation s . By dividing each average by its corresponding standard error, variation across replicates can be accounted. The mean expression is modified to the standardized mean expression levels, (M_i) and $(t_i) = (M_i / SE_i)$, where $(SE_i) = s_i / \sqrt{n}$ is the standard error (SE) of M_i , s_i being the standard deviation of the values $M_{ij}, j = 1, \dots, n$. A standard (Student) t-statistic is calculated as given in 2:

$$t = \frac{\bar{M}}{(s / \sqrt{n})} \quad (2)$$

t : t – value

\bar{M} : median M – value (Rank)

s : standard deviation

n : sample size

M is the log transformed ratio from two samples, s is the standard deviation and n the scope of random sample inspection. From this formula (2) it is clear that a large t -statistic (and the corresponding highly significant p -value), can occur because of either a large M (high ratio) or small s (low noise).

High t -values lead to small p -values. A result of this test (2) is significant, if the calculated p -value is smaller than the previously defined level of significance. The p -value of an observed value t_{observed} of some random variable T used as a test statistic (2), is the probability for a true null hypothesis, meaning T will assume a value less or greater to the null hypothesis as the observed value t_{observed} . A statistical test reflects the probability that a null hypothesis is true. The lower the value the less likely the null hypothesis is true, which states there is no difference. Assumptions sorted by p -value can be sorted by their significance. The lower the p -value, the more significant are e.g. the genes differentially regulated among these conditions. In other words, the p -value is the probability of obtaining a similar result for a given data point, assuming the data point was just the result of chance. The disadvantage of this method is that genes with a very (sometimes artificially, due to random chance) small variance will obtain high t -statistic-values, although the genes might not be differentially expressed. A lot of false positive still prevent the standard t -statistic from serving as a reliable test of which genes are truly regulated.

To overcome these limitations it is possible to introduce a correction factor. Various modifications to the t -statistics have been proposed. A more sophisticated statistical method refining t was introduced (Tusher *et al.*, 2001). Adding a constant term " a " to the denominator of the standardized average prevents the dominator from getting too small for low s (3). This reduces the number of false positive genes at unusually low s . Several alternative statistics (e.g. 4) have been proposed to overcome these problems by

using the "penalized" t-statistic (also called "moderate" or "regulated" t-statistic) (Efron *et al.*, 2000). The calculation of this penalized t-statistic is as follows:

Penalized t-statistic (Tusher):

$$t = \frac{\bar{M}}{\sqrt{(a + s)^2 / n}} \quad (3)$$

t : t – value

\bar{M} : median *M* – value (Rank)

s : standard deviation

n : sample size

a : penalty

Penalized t-statistic (Efron):

$$t = \frac{\bar{M}}{(a + s) / \sqrt{n}} \quad (4)$$

t : t – value

\bar{M} : median *M* – value (Rank)

s : standard deviation

n : sample size

a : penalty

Choosing too large "*a*" values makes the denominator a constant, resulting in a loss of information about the variability of the measurements. Therefore estimating the optimal "*a*" for a particular dataset is crucial, often the 90th percentile of the *s* values is chosen. Ranking genes from a microarray dataset, a penalized t-statistic can perform better than a standard (Student) t-statistic in terms of decreasing the false positive and false negative rate (Comander *et al.*, 2004). Using this modified t-statistics approach, implemented in the gene-data software package "Analyst", up- and down-regulated genes were determined. Further data mining, data transformation, visualization and modelling was done using scatter plots, principle component analysis, pathway analysis and cluster analysis as shown in the results part.

A scatter plot is two dimensional graphical representation of expression data in which the intensity of two samples are plotted on the x- and y-axis (Figure 23).

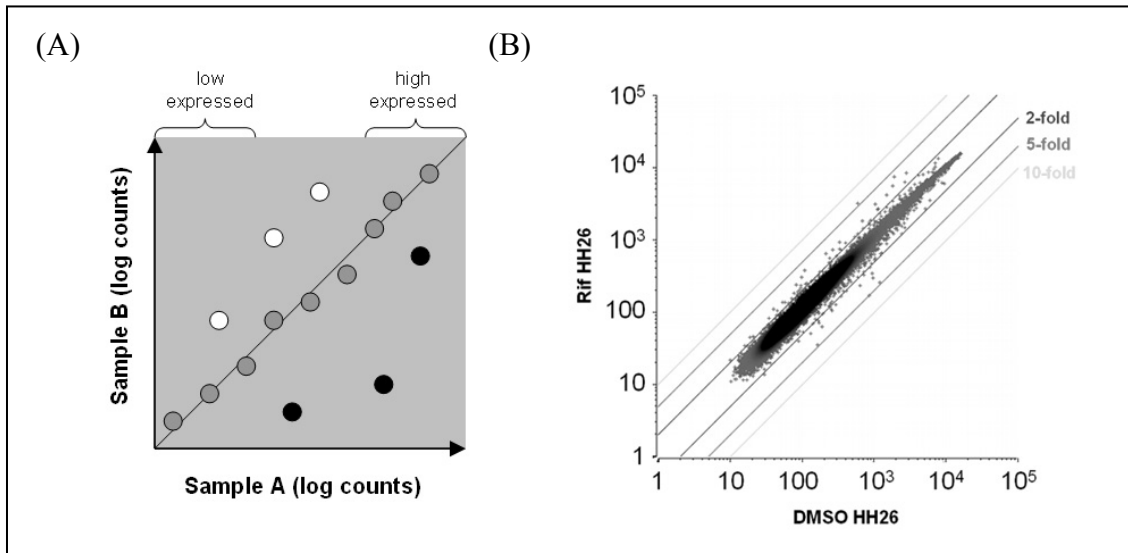


Figure 23: A useful representation of expression data: the scatter plot. (A) A schematic drawing of a scatter plot showing 15 genes (circles) expressed in two samples (A and B). Nine of the genes show a similar expression (grey circles on the diagonal). Three genes (white) are abundant transcripts, induced in sample B relative to sample A. Three genes (black) are repressed in sample B compared to sample A. (B) Scatter plot from this work representing 30,000 genes from a full genome array. The lines indicate a two-, five-, and ten-fold different expression. Genes that are found above the lines are up-regulated in the Rifampicin ("Rif") treated samples compared to the control sample (DMSO) while genes below the lines are down-regulated.

The principle component analysis (PCA) is a multivariate technique which allows the comparison of many different variables (genes). It preserves close relationships between genes in a multi dimensional space. This means that genes in close proximity in many dimensions are also close to each other in the three dimensional representation (Figure 24).

Principle component analysis is used to identify groups of genes that are regulated in a similar manner across many experiments and conditions. In the experiments presented here the experimental conditions were: n patients ($n = 3$) and p criteria (two treatments and three time points). To visualize the data, n -spots had to be plotted in a p -dimensional space R^p , which is graphically impossible ($p > 3$). PCA projects these spots in a q -dimensional subspace R^q ($q < p$). The variables (patient, treatment and time) are shown in a lower number of linear combinations of principle components, which allows visualization of groups (patients = arrays) showing a similar expression behaviour.

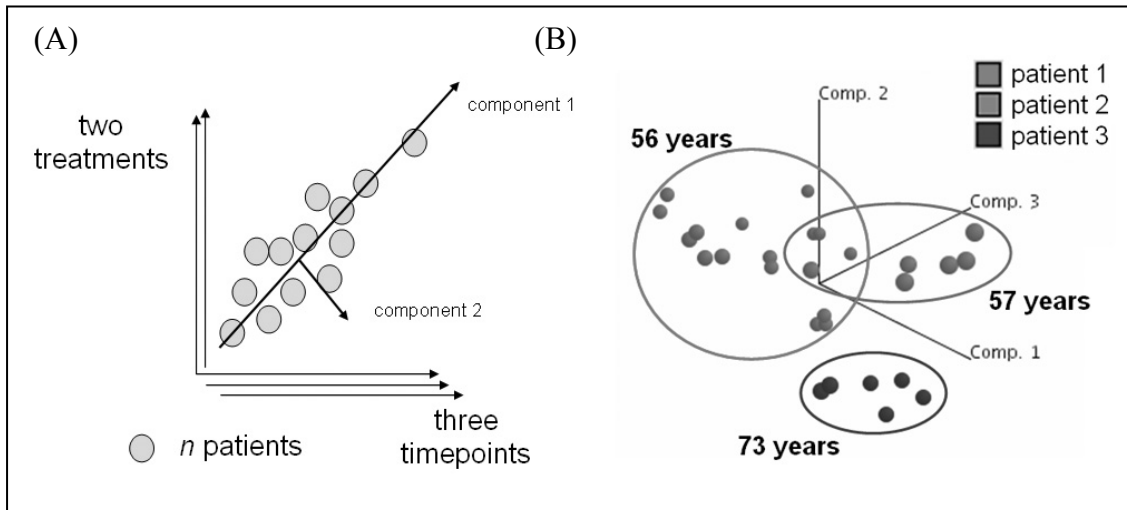


Figure 24: Principle component analysis reduces relationships that exist in a high dimensional space into only three dimensions, enabling complex data to be visualized. Part (A) shows the basic principle of the computational method of PCA. A new coordinate system is applied into the point cloud of observations such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component; variance is a degree of containing information). The second axis with the second highest variance is placed orthogonal on the first followed by the third which is placed orthogonal on the previous two. The content of this new axis (components) could often not be interpreted, but specific patterns will show up as indicated in (B). Part (B) gives an example of a PCA from this work and it can clearly be seen that over all patients, treatments and time points, experiments (= arrays) with the same or similar age of the patients from which the samples were derived, fall together.

Expression data can further be superimposed onto biochemical and genetic pathways to correlate gene expression with metabolic activity (DeRisi *et al.*, 1997). This computational method of pathway analysis correlates gene expression values with biochemical, genetic and developmental processes and gives a deeper insight in the functional meaning of the array data. This mapping approaches rely on previous knowledge as it is available for well known biological pathways like central carbon metabolism (glycolysis, tricarboxylic acid (TCA) cycle, amino acid- and nucleotide metabolism and synthesis, and so on). For the metabolic and especially for the regulatory network of hepatocytes it is very difficult as no complete networks exist. The presented work here will help to improve this situation and propose at least some sub-networks of gene regulation in human hepatocytes.

Another multivariate classification method is cluster analysis. Using this technique data sets are grouped based on their similarity. Similar data is collected in one cluster. Clusters are organized in a way that similar clusters are close to each other and different clusters are far apart. Clustering was first applied to microarray analysis in the late 1990s (Eisen *et al.*, 1998). A typical gene expression clustering diagram displays the genes and gene clusters vertically and each experiment is given on the horizontal axis. The ratios are often colour coded, so that activated genes appear in red and repressed genes are displayed in green. The tight correlation between gene expression and function means that genes that cluster together often share a common function. Thus clustering provides a rapid method to identify and visualize genes that share a similar function or at least expression behaviour.

All techniques and methods given here are used to describe, analyse and visualize the expression data gained from the microarray experiments to identify differentially expressed genes and enabling further functional annotation and interpretation.

1.5 Objective of this Thesis

Gene expression profiling is one of the most powerful and informative analysis methods of biological function. Complex biological processes are currently best studied by systematic and holistic monitoring of gene expression patterns. Gene expression strongly correlates with cellular function, therefore gene expression data is widely used as the source for information on biological systems and basis for mathematical models. As model system for the human liver the gene expression response of primary human hepatocyte cell cultures, comparing Rifampicin treated and non-treated samples, was monitored.

The liver was chosen because this organ

- plays a major role in endogenic metabolism
- plays a central role in the metabolism of xenobiotics
- has as an organ the unique ability to regenerate

The investigation of regulatory and metabolic networks in the liver is of high interest to relevant medical and industrial questions.

For understanding, modelling and further investigating this complex organ a systems biology approach was launched obtaining the transcriptome, proteome and metabolome of primary human hepatocytes. The work presented here focuses on time resolved analysis of gene expression after perturbation of liver cells with Rifampicin. Rifampicin, which is a synthetic bacterial antibiotic drug is typically used to treat *Mycobacterium* infections, including tuberculosis and leprosy and plays also an important role in the treatment of methicillin-resistant *Staphylococcus aureus*. Rifampicin was chosen as a model xenobiotic, as it is a potent inducer of hepatic cytochrome P450 enzymes (such as CYP2D6 and CYP3A4) and will increase the metabolism of bile acids (endogenous metabolism). Many drugs are transformed and prepared for excretion by the liver through this enzyme system (xenobiotic metabolism). Therefore this approach represents a versatile setup for investigating liver

regulation and function. The objective is to obtain a broader insight in the regulation and metabolic principles leading to high variability, redundancy and stability of liver functions.

The limited availability of human hepatocytes results in the use of animal cells in a lot of liver expression studies. It is however important to realize that humans differ from animals with regard to isoform composition, expression and catalytic activities of drug metabolizing enzymes (Guengerich, 1997; Bogaards *et al.*, 2000). Especially CYP1A, CYP2C, CYP2D, and CYP3A show interspecies differences in terms of catalytic activity and regulation (Martignoni, 2006). The focus of this work was the generation of data for assembling a model of the human liver. As mentioned, caution should be applied when extrapolating metabolism data from animal models to human. Although it is quite challenging to obtain the necessary amount of usable human cells compared to mouse or rat, the analysis was performed with primary human hepatocytes to be as close as possible to the reality found in humans.

Hepatocytes were cultivated and deflected by administering Rifampicin. At several time points, cells were harvested and the gene expression level of the cells was analysed by microarrays and real-time PCR (qRT-PCR). Gained data are used to generate several hypotheses on interactions of the induced pathways and enzymes maintaining liver function (like detoxification and maintaining glucose homeostasis). In cooperation with the Institute of Biochemical Engineering, University of Stuttgart (Prof. Reuss and P. K. Murugan) these data are the basis for dynamic modelling of the regulatory network in response to a drug perturbation in human hepatocytes.

It is known that xenobiotics, including drugs, can influence cytochrome P450 (CYP) activity by up regulating the transcription of CYP genes. However the underlying molecular mechanisms remain partially unknown. To minimize potential drug interactions, it is important to ascertain if a compound will be an inducer of CYP enzymes. Therefore a deeper insight in the regulatory circuits is crucial for understanding and maybe even prediction of induction of metabolic enzymes. So far this had to be examined experimentally using animals or hepatocyte cell cultures. The presented work is a starting point for the model driven *in-silico* prediction of enzyme (CYP) regulation in the human liver.

2 Materials and Methods

2.1 Chemicals, Enzymes and Kits

Basic chemical compounds were used from the companies *Fluka*, *Merck*, *Riedel-deHaen* and *Roth* in analysis grade or higher. Enzymes were from the companies *Ambion*, Palo Alto, *MBI Fermentas*, St. Leon Rot, *Invitrogen*, Carlsbad and *Eppendorf*, Hamburg. Ampicillin was used from the company *Boehringer Mannheim GmbH*, Mannheim. Trizol was bought from *Invitrogen*, Carlsbad. Yeast extract and Nutrient Broth was from *Difco Laboratories*, Detroit, USA, Trypton from *Oxoid*, Wesel. The used agarose was from the company *Pharmacia Biotech GmbH*, Freiburg. From *Serva Feinchemikalien GmbH*, Heidelberg bromophenol blue, glycine, Lysozyme, urea, Na₂-EDTA, SDS, TEMED and TRIS were used.

For the extraction of nucleic acids the following kits were used: purification of RNA was done with RNeasy[®] Mini Kit and for the isolation of plasmid DNA the Genelute[™] Plasmid Mini Prep Kit from *Sigma* was used.

Amplification of RNA was done with the MessageAmp[™] II aRNA Kit from *Ambion*. Furthermore the Super Signal West Dura Kit from *Pierce* was used for luminescence detection of western blots.

2.2 Hardware Equipment

Equipment used in this work is shown in the following table (Table 1):

Table 1: Laboratory equipment and instruments used in this work

instruments and hardware equipment	type	manufacturer / company
capillary gel electrophoresis	Agilent 2100 Bioanalyzer	Agilent Technologies, Waldbronn, Germany GmbH
agarose gel electrophoresis device	Mini-Sub™ DNA Cell	BioRad, München, Germany
power supply for gel electrophoresis	Mini-Sub™ DNA Cell GT	
western blot device	BioRad Power PAC 3000 / 300	BioRad, München, Germany
centrifuge	Trans-Blot SD SEMI DRY Transfer Cell	BioRad, München, Germany
vacuum concentrator ("speedvac")	Sorvall RC-5B	Du Pont Instruments, Bad Homburg, Germany
table centrifuge	Concentrator 5301	Eppendorf, Hamburg, Germany
table shaker ("vortexer")	Centrifuge 541 7C Centrifuge 541 7R	Eppendorf, Hamburg, Germany
heated magnetic stirrer	Janke und Kunkel Vortex Genie 2	IKA Labortechnik, Stauffen, Germany
microwave	RCT Basic	Scientific Industries, Bohemia, USA
UV light table	Micro-Chef FM A935	IKA Labortechnik, Stauffen, Germany
photometer Nanodrop	UV light table	Moulinex, Ecully Cedex, France
balances	ND-1000 Spectrophotometer	MWG Biotech, Ebersberg, Germany
microarray scanner	Precision Advanced Basic	Nanodrop Technologies, Delaware, USA
pH meter	ScanArray Express Microarray Scanner	Ohaus Waagen, Giessen, Sartorius, Göttingen, Germany
Incubators	Digital pH Meter pH525	PerkinElmer, Kalifornien, USA
quantitative real-time PCR	HT incubator multitron	WTW, Weilheim, Germany
quantitative real-time PCR	ABI Prism™ 7500 Sequence Detector	Infors AG, Switzerland
	Mastercycler realplex ² epgradient S	Applied Biosystems, Foster City, CA U.S.A.
		Eppendorf, Hamburg, Germany

2.3 Software Tools and Internet Resources

Besides the common used software for picture editing, text- and data processing (CorelDraw, Microsoft Office, Photoshop, and other), the following special applications were used (Table 2):

Table 2: Software applications used in this work

software	company / author	application
R Version 2.1.1	R Development Core Team	platform for the application of limma, sma and statmod (for the statistical analysis of DNA-microarrays)
ScanArray Express Version 3.0	PerkinElmer	scanning and quantification of DNA-microarrays
DualChip® Evaluation Software 1.1	Eppendorf	evaluation of Eppendorf DualChips®
2100 expert Version B.02.02.SI238	Agilent	evaluation (determination of RIN) of capillary electrophoresis chips (Bioanalyzer)
WinSCP3 Version 3.7.6 (Build 306)	Free Software Foundation, Inc.	file transfer program (data transfer and exchange)
MDL ISIS™/Draw 2.5	MDL Information Systems, Inc.	drawing of chemical structures
Refiner	Genedata	statistical analysis and pre-processing of DNA-microarrays
Analyst	Genedata	data evaluation of (Affymetrix) DNA-microarrays
Philosopher	Genedata	biological data base for interpretation of omics data (e.g. DNA-microarray data)
Bibliosphere	Genomatix	LMMA (literature mining and array analysis)

The following references show web pages used in this work (Table 3):

Table 3: Web pages used in this work

address	application
https://www.affymetrix.com/analysis/netaffx/index.affx	annotation of Affymetrix probes
http://www.r-project.org/	software for statistical analysis
http://stat-www.berkeley.edu/users/terry/zarray/Software/smacode.html	R package: statistics for microarray analysis
http://www.bioconductor.org/	software project with free source code for analysis of genome data
http://www.ncbi.nlm.nih.gov/	database for molecular biological information
http://sysbio.mpi-magdeburg.mpg.de/	Hepatosys Magdeburg exchange server
http://www.biobase-international.com/pages/	biological database for e.g. pathway analysis
http://www.montp.inserm.fr/u632	Institut National de la Santé Et de la Recherche Médicale

2.4 Buffers and Solutions

The following media and antibiotics for cultivation of microorganisms were used in the work presented here. For solid media 1.5 % (w/v) agar was added to liquid culture media. This solution was autoclaved immediately after preparation for 20 min at 121°C.

2.4.1 Complex media

Luria-Bertani-(LB-)Media Sambrook *et al.*, 1989

Trypton	10 g
Yeast Extract	5 g
NaCl	10 g
H ₂ O _{bidest}	ad 1000 ml

pH 7.5 (NaOH)

2.4.2 Antibiotics

Antibiotic stock solutions were prepared as given in Table 4 based on Sambrook *et al.*, 1989, divided into aliquots and frozen at minus 20°C. The final concentration of Ampicillin (sodium salt) was 100 µg/ml and was diluted from a stock solution of 100 mg/ml. Ampicillin stock solution was added after the liquid media was cooled down below 50°C.

Table 4: Stock solutions and working concentrations of common antibiotics based on Sambrook et al., 1989

	stock solution		working concentration	
	concentration	storage	stringent plasmids	relaxed plasmids
Ampicillin	50 mg/ml in H ₂ O	-20°C	20 µg/ml	60 µg/ml
Carbenicillin	50 mg/ml in H ₂ O	-20°C	20 µg/ml	60 µg/ml
Chloramphenicol	34 mg/ml in EtOH	-20°C	25 µg/ml	170 µg/ml
Kanamycin	10 mg/ml in H ₂ O	-20°C	10 µg/ml	50 µg/ml
Streptomycin	10 mg/ml in H ₂ O	-20°C	10 µg/ml	50 µg/ml
Tetracycline	5 mg/ml in EtOH	-20°C	10 µg/ml	50 µg/ml

^a Stock solutions of antibiotics dissolved in H₂O should be sterilized by filtration through a 0.2-micron filter. Antibiotics dissolved in ethanol need not be sterilized. Store solutions in light-tight containers.

^b Magnesium ions are antagonists of Tetracycline. Use media without magnesium salts (e.g., LB medium) for selection of bacteria resistant to Tetracycline.

2.5 Media for Cultivation of Primary Human Hepatocytes

The composition of media for the cultivation of primary human hepatocytes is given in the following tables (Table 5 and Table 6):

Table 5: Composition of the culture medium for the cultivation of human hepatocytes

additives	concentration	ad 500 ml WME
fetal calve serum	10 % (v/v)	50 ml
penicillin/streptomycin	100 U/ml; 100 µg/ml	5 ml
dexamethasone	0.025 % (v/v)	125 µl
insulin	0.032 IU/ml	400 µl

Table 6: Culture medium components for cultivation of human hepatocyte cells. Several components were obtained from the Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart and University Tübingen indicated by (RBK)

medium / additives	company	order number/ PZN number
Williams Medium E with stable Glutamine, 2.24g/l NaHCO ₃	PAN Biotech GmbH	P04-29150
fetal calve serum	Biochrom KG	S 0113
penicillin (10,000 U/ml)/ Streptomycin (10,000 µg/ml)	Gibco / Invitrogen	4032
Insuman Rapid 40 IU/ml (1.4 mg/ml insulin human) injection solution	Aventis (RBK)	PZN 1843315
Dexa Inject (8 mg Dexamethasone- dihydrogenphosphate/ 2 ml injection solution)	Jenapharm (RBK)	PZN 8704396

2.6 Primary Human Hepatocyte Cells, Bacterial Strains and Plasmids Used in this Work

Hepatocyte cells used for these studies were obtained from surgeries, taken into culture and were delivered by the platform cell biology within the Hepatosys network (mainly from Prof. Nüssler). An overview of the samples (no., size, delivery format and supplier) and additional patient specific information concerning age, sex and diagnosis are given in Table 7 and Table 8. Furthermore the experiments performed are summarized, as well as the array format used to investigate the expression levels in these cells.

Table 7 gives a summary for the full genome approach using Affymetrix arrays. It can be seen that all patients were male. Patient number HH26 (HH = human hepatocytes) was 56 years old and had a hepatocellular carcinoma. The second patient HH27 was 58 and no diagnosis was obtained for this patient. The third patient HH44 was 73 years old and suffered from metachrome liver metastases. For all three patients the same experimental setup was used. Cells were split into two groups, one treated with Rifampicin and the control with DMSO. At 6, 24 and 72 h samples were collected and frozen in liquid nitrogen for further processing. For the culture HH26 additional induction experiments with Phenobarbital (PB) were performed and cells were harvested at the same time points. Furthermore untreated cells were collected 24 h

before induction of the other cells, at the point of induction as well as after 72 h to investigate any change based on dedifferentiation of cells in culture.

Table 7: Used samples for transcriptomics and proteomics profiling. The table shows patient specific data as sex, age and diagnosis, gives the total number of received cells, the delivered format (6 well plates or T75 culture bottles) and the supplier. Finally it presents an overview which experiments were performed using Affymetrix full genome chips.

culture no.	date	supplier	delivery format	total amount of cells	sex	age	diagnosis	experiments	micro-array format
HH26	9.12.05	Prof. Nüssler Berlin	6x6-well 6xT75 bottle	300 mio	<i>m</i>	56	hepatocell. carcinoma (HCC)	Transcriptomics: Time series (DMSO, PB, RIFA) ; RNA after 6h, 24h, 72h dedif. -24h, 0h, 72h. RNA Proteomics: Time series (DMSO, RIFA), 6h, 24h, 72h	Affymetrix
HH27	16.1.06	Prof. Nüssler Berlin	16 x T75 bottle (6 for IKP)	300 mio	<i>m</i>	58	n.d.	Transcriptomics: Time series (DMSO, PB, RIFA) ; RNA after 0 h, 6h, 24h, 72h	Affymetrix
HH44	24.8.06	Prof. Nüssler Berlin	6 T 75 bottles 4 x 6 well	126 mio	<i>m</i>	73	metachrome liver metastases	Transcriptomics: Time series (DMSO, RIFA) RNA after 6h, 24h, 72h Proteomics: Time series Induction 6h, 24h, 72h. (hypoosmotic HEPES-buffer + protease inhibitors)	Affymetrix

Cell culture from patient HH27 was delivered in T75 culture bottles so the experiment had to be done in this larger volume of cells instead of six wells plates. However induction conditions and points for collecting samples (6, 24 and 72 h after induction) were kept the same as for sample HH26. HH44 was again performed in six-well plates with the same experimental conditions as for the previous two samples.

Table 8 gives the same information for samples used in sub-genome expression profiling experiments using Eppendorf hepato DualChips. These chips, as they cover only a small subset of detoxification relevant genes (151), are comparable cheap and therefore allowed analysis of more samples with a higher sampling frequency in a time course experiment. As shown in Table 8 the induction of cells from sample HH38 by Rifampicin was analysed after 2, 8, 16, 24, 32 and 63 h in six well plates. The amount of cells from HH43 for transcriptomics analysis only allowed the investigation of three time points. The expression level of genes was monitored for 6, 24 and 72 h.

Table 8: Samples used for profiling the transcriptome and proteome. The table shows patient specific data as sex, age and diagnosis, gives the total number of received cells, the delivered format (6 well plates or T75 culture bottles) and the supplier. Finally it presents an overview of which experiments were performed using Eppendorf sub-genome hepato Dual arrays.

culture no.	date	supplier	delivery format	total amount of cells	sex	age	diagnosis	experiments	micro-array format
HH38	31.5.06	Prof. Nüssler Berlin	6x6-well	54 mio	<i>m</i>	70	colorectal liver metastases	Transcriptomics: Times series (DMSO, RIFA) RNA after 2h, 8h, 16h, 24h, 32h, 63h Proteomics: Induction (DMSO, RIFA) 40h (AQUA-Analyse)	Eppendorf
HH43	17.8.06	Prof. Nüssler Berlin	6 T75 bottles 2x 6-well	108 mio	<i>f</i>	83	hepatocell. carcinoma (HCC)	Transcriptomics: Time series (DMSO, RIFA) RNA after 6h, 24h, 72h Proteomics: Time series Induction 6h, 24h, 72h. (hypoosmotic HEPES-buffer + protease inhibitors)	Eppendorf

For cell culture HH29 and HH48 no patient data was received.

Table 9: Sample of human hepatocytes (HH) for time series experiments analysed with real-time PCR. No patient specific data was obtained for this sample. The table provides the total number of cells received, the format delivered (6 well plates) and the supplier. Finally it presents an overview of the experiments performed using real-time PCR and western blot.

culture no.	date	supplier	delivery format	total amount of cells	sex	age	diagnosis	experiments	format
HH29	8.3.06	AG Nüssler Berlin	10x6-well	90 Mio	-	-	-	Transcriptomics: time series (DMSO, Rifampicin) 0.5, 1, 4, 8, 12, 16, 24, 30, 36, 42, 48h. RNA (TaqMan) westernblot-Analyse	TaqMan

As standards and positive controls for real-time polymerase chain reaction (qRT-PCR) the following bacterial strains and plasmids (Table 10) were used to generate plasmid DNA containing the suitable target gene:

Table 10: Bacterial strains and plasmids used in this work

strain / plasmid	geno- / phenotype	reference
<i>Escherichia coli</i> DH5 α	endA1, recA1, dcoR, hsdR17 (rk-mk-, supE44, thi-1, λ -, gyrA96, relA1, Δ (lacZYA-argF)U169 [ϕ 80dlacd(lacZ) M15]	Invitrogen/Gibco BRL, Rockville, Md.)
pUVI CYP3A4	Plasmid approximately 12 kB CYP3A4 positive control plasmid for qRT-PCR	Dr. U. Zanger (Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart and University Tübingen)
CYP3A4	Plasmid 6 kB CYP3A4 positive control plasmid for qRT-PCR	Dr. O. Burk (Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart and University Tübingen)

2.6.1 Cultivation of Bacterial Cells

Cultivation of *E. coli* strains was usually done in 2 to 10 ml (preparatory culture) and 50 ml (main culture) in a shaking flask with baffles. The ratio between liquid- and gas phase was always at least 1:10 to assure an optimal oxygen supply to the cells. The cultures were either inoculated with a single cell from a freshly grown plate or with 0.1 volume (10 %) of a densely grown preparatory culture. The incubation of the cultures took place at 37°C and 180 rpm in a shaking incubator from Infors (Infors, Bottmingen, Switzerland).

The growth of the cells could be followed by photometric measuring the optical density at a wavelength of 600 nm in a spectral photometer (BioPhotometer, Eppendorf).

2.6.2 Conservation of Strains

For short periods (e.g. *E. coli*, up to one week) bacterial strains were stored at 4°C on suitable agar plates which were sealed with Parafilm.

Long term storage of *E. coli* strains was done at - 80°C as glycerol culture (15 % v/v glycerol 170 µl 87 % glycerol + 830 µl cell culture). To reactivate the frozen cells they were plated directly from the stock on suitable LB agar plates.

2.7 Cell Cultivation and Induction of Primary Hepatocytes

Recovery of high quality RNA from primary human hepatocyte cells at several different points in time required cell cultivation in house. Therefore obtained cells were cultivated in the cell laboratory in collaboration with the Institute of Biochemical Engineering, University of Stuttgart. The experimental procedure of induction consisted of three parts. The first step was recovery and documentation of cells after transport. After usually 1 day of resting the cells were used for induction experiments and finally cells were harvested and total RNA was isolated.

2.7.1 Bringing Primary Hepatocytes into Culture after Transport

Cells were disseminated on collagen coated six well plates. After delivery, the cells were immediately incubated in fresh culture medium. Under sterile conditions the sealing foil was removed and a new cover was applied to the plates. The old culture medium was removed using a disposable pipette and a vacuum pump. Subsequently to every well 2 ml of 37°C pre-warmed culture medium (see 2.5) was carefully added to the adhesively growing cells minimizing any perturbation. The date, time of arrival and noticeable distinctive features like unusual density of growth, contamination, and detachment were recorded. Finally a photo was taken to document the morphology and actual condition of the cells after arrival. To let the cells regenerate they were incubated over night at 37°C and 5 % CO₂. Till the beginning of the time series experiments the change of medium was done every 24 h.

2.7.2 Induction Experiments of Primary Hepatocytes

Differential gene expression over time was analysed using the following experimental setup. After induction of primary human hepatocytes with a drug (e.g. phenobarbital (PB) or Rifampicin (Rif) solved in DMSO) samples were taken after three points in time for microarray and qRT PCR analysis. In parallel control samples of the same culture were isolated, which were only incubated with DMSO to take possible effects of the solvent into account. For each treatment and point in time approximately 1.5 million cells were used. For a complete analysis every sample was processed as duplicate, so altogether 6 million cells for each particular point in time, control- as well as treated sample were used (two times treated, cells with DMSO and drug and two times control, cells incubated with DMSO only). The following table (Table 11) gives a scheme of the experimental setup:

Table 11: Schematic overview of an induction experiment with primary hepatocytes. DMSO-c stands for a sample only incubated with DMSO (control). 30 μ M Rif indicates a sample induced by Rifampicin. Every 24 h the culture medium was changed. The fresh medium contained the same concentration of drug and solvent as for the induction to maintain the induction.

day 2 after cell isolation	day 3 after cell isolation	day 4 after cell isolation	day 5 after cell isolation	day 6 after cell isolation
arrival	start of	change of	change of	
change of medium	induction	medium/ 2nd induction	medium/ 3rd induction	
2 x 6 well plates	1. DMSO 2. 30 μ M Rif			
	6 h, 24 h, 72 h	72 h	72 h	
	harvesting 6 h:	harvesting 24 h:		harvesting 72 h:
	DMSO 30 μ M RIF	DMSO 30 μ M RIF		DMSO 30 μ M RIF
	(2 wells, 3 mio cells/ treatment)	(2 wells, 3 mio cells/ treatment)		(2 wells, 3 mio cells/ treatment)

The experiments were started two days after isolation of the cells and one more day of regeneration in the cell culture laboratory in Stuttgart. The cells were cultivated in 2 ml WME complete medium / well (WME + GlutaMax + 10 % FCS, 1 % Pen/Strep, 0.032 IU/ml Insulin, 0.025 % Dexamethasone, see 2.5).

Depending on the type of sample (control, induced) different kinds of media were added to the samples: 0.1 % DMSO for the control sample (control), 500 μ M Phenobarbital (from a 200 mM stock solution) + 0.1 % DMSO (induced PB) respectively 30 μ M Rifampicin (from a 30 mM stock solution, solved in DMSO equal to 0.1 % DMSO) (induced Rif). Every 24 h the medium containing the inducer and solvent (as indicated above) was changed.

2.7.3 Harvesting of Primary Hepatocytes

Cells were washed two times in sterile PBS pre cooled to 4°C. Following this step 350 μ l RLT-Puffer (Guanidinisoithiocyanat Puffer) + Mercaptoethanol (10 μ l β -ME in 1 ml RLT-Puffer) was added to every well ($< 5 \times 10^6$ cells). Subsequent incubation step, led to cell lysis. Cell lysate was scraped with plastic cell scraper and homogenized by pipetting the suspension several times up and down in a 1 ml pipette (as long as a noticeable reduction of the viscosity could be observed). The lysate from every well was separately applied to a Qiagen-Shredder column to improve homogenisation. The samples were centrifuged for 2 minutes at 14000 rpm. The tubes were sealed with a cap and labeled with a sticker indicating date and time of preparation and kind of sample. The lysate was frozen in liquid nitrogen and stored at - 80°C for further analysis.

2.8 Methods of Molecular Biology

2.8.1 Working with RNA

Working with RNA needs some precautions which prevent degradation of RNA which is more sensitive especially under alkaline conditions (oxygen will be more nucleophilic) compared to DNA. This instability is based on the additional hydroxy group in the 2'-position of ribose. In addition Ribonucleases (RNases) are very stable and ubiquitous active enzymes that generally do not require cofactors to function. Since RNases are difficult to inactivate and even very small amounts are sufficient to destroy RNA, all laboratory equipment used must be prepared to prevent possible RNase contamination. For non-disposable plastic ware this was achieved by rinsing with 0.1 M NaOH, 1 mM EDTA followed by RNase-free water. Alternatively, chloroform-resistant plastic ware can be rinsed with chloroform to inactivate RNases. Glassware was cleaned with a detergent (e.g. 0.5 % SDS), thoroughly rinsed and heated to 240°C for four or more hours before use. Alternatively, glassware was treated with DEPC (diethyl pyrocarbonate) or RNase Zap from Ambion. Glassware was filled with 0.1 % DEPC (0.1% in water), and incubated over night at 37°C. To eliminate residual DEPC the glassware was autoclaved or heated to 100°C for 15 min after incubation. The use of sterile, disposable polypropylene tubes does not require any precautions. These tubes are generally RNase-free and do not require pre-treatment to inactivate RNases.

Solutions (water and other solutions) were treated with 0.1 % DEPC. DEPC is a strong, but not absolute inhibitor of RNases. It inactivates RNases by covalent modification, as it reacts with primary amines. Therefore it can not be used to treat TRIS buffers. 0.1 ml DEPC was added to 100 ml of a solution to be treated and vigorously mixed. The solution was incubated for 12 hours at 37°C. Finally it was autoclaved for 15 minutes to remove any trace of DEPC.

Furthermore proper microbiological, aseptic technique was used when working with RNA. Hands and dust particles may carry bacteria and moulds and are the most common sources of RNase contamination. Therefore it is advisable to always wear latex or vinyl gloves while handling reagents and RNA samples to prevent RNase contamination from the skin.

2.8.2 Purification of Nucleic Acids

Total RNA from hepatocytes were isolated and purified following the instructions of the "RNeasy[®] Mini Handbook" (Third edition, June 2001) from the company Qiagen. This procedure allows a quick isolation of up to 100 µg total RNA from primary cell culture. The basic principle behind it is a specialized high-salt buffer system which assures a selective binding of RNA at the silica-gel-matrix.

Cells were denatured in a strong guanidinisothiocyanate buffer (GITC) (see 2.7.3). This buffer is lysing the cells and additionally inactivates RNases and therefore allows isolation of not degraded RNA. The lysate was frozen and stored at -80°C. For subsequent processing the frozen lysate was heated for 15 minutes to 37°C and mixed well. One volume of Ethanol was added to the thawed lysate to adjust optimal binding conditions. The solution was well mixed by pipetting and applied to an RNeasy mini column, which contained the silica-gel-matrix. The RNeasy mini column, placed in a 2 ml collection tube was centrifuged for 15 sec at 8000 g. The flow-through was transferred to a new tube for further protein analysis like SDS-PAGE and western blotting and put aside. All RNA molecules longer than 200 nucleotides are isolated. The procedure provides enrichment for mRNA since most RNAs < 200 nucleotides, such as 5.8S rRNA, 5S rRNA and tRNAs, which together comprise 15-20 % of total RNA, are selectively excluded.

Contaminations with genomic DNA may lead to wrong results in subsequent analysis like microarray experiments or real-time PCR. Therefore residual DNA was removed using the RNase-free DNase set for optional on-column DNase digestion. After purification the bound RNA was eluted with 50 µl of RNase free water from the column. To obtain a higher total RNA concentration the elution step was repeated using the first eluate. If necessary the eluate could be further concentrated by precipitation (see 2.8.4) leading to a higher concentration and purity or using a vacuum concentrator.

2.8.3 Quality Control of Nucleic Acids

The quality of isolated total RNA was controlled by photometric measurements, formaldehyde-gel-electrophoresis and capillary gel-electrophoresis using the Agilent Bioanalyzer 2100 (RNA 6000 Nano LabChip Kit).

The photometric analysis was done with the NanoDrop. Using this spectrophotometer it is possible to record the nucleic acid spectrum (221 to 453 nm) with only 1 μ l of sample. The associated software (ND-1000 V 3.1.2) allows the visualisation of the spectrum and calculates the concentration based on the absorption at 260 nm. The absorption ratios between 260 nm / 280 nm and 260 nm / 240 nm which are a measure for purity are also calculated. For pure DNA the ratio is 1.8, while a ratio of 2.0 indicates pure RNA. Ratios smaller than the mentioned values, are a clear hint for the concomitance of proteins, phenol, residual washing buffer or other contaminants in the sample. The second value to assess purity is the absorption ratio at 260 nm / 230 nm. The values for this quotient for pure nucleic acids should be higher than the associated 260/280 value. Generally these values range from 1.8 to 2.2. Significantly lower values are a sign for contaminants in sample.

Furthermore based on the Nanodrop spectrophotometer measurements it was possible to calculate the incorporation rate of fluorescent dyes which are used for labeling cDNA during the reverse transcription reaction. To be able to calculate the amount of labeling the concentration of nucleic acid as well as the fluorescence of the respective dye was measured. A sample calculation is given in results part (3.5.1).

The quality concerning integrity of RNA was determined by denaturing 1% formaldehyde-gel-electrophoresis (see 2.8.7). RNA integrity is assessed by visualization of intact ribosomal RNA bands. For the total RNA from higher eukaryotes, the size of the ribosomal bands should be 1.9 kb for the 18S-RNA and 4.7 kb for the 28S-RNA. Both rRNA bands (18S and 28S) should be fairly sharp and intense. The upper 28S band should be twice as bright as the lower 18S band. If degradation of RNA occurred a smear below the 28S band is visible in the gel and the intensity of this band is significantly reduced.

Another method to determine concentration and quality of RNA samples is the electrophoretic separation using capillary gel-electrophoresis in a RNA 6000 Nano Chip of the Agilent Bioanalyzer. For this analysis only 1 μ l of RNA is needed. During

the run the sample is labeled with a fluorescent dye, separated and analysed using a microfluidic chip. Next to the twelve samples which are analysed in parallel, an internal standard for normalization (artificial nucleic acid with known size) and a molecular weight marker (RNA 6000 ladder from Ambion) are analysed. The ladder contains 150 ng/ μ l RNA and allows a comparison in size and a relative quantification of the unknown samples.

The experimental procedure was done according to the instructions of the supplier. As result an electropherogramm (which shows the fluorescence intensity over time), the concentration of the sample and the ratio of the ribosomal peaks (28S to 18S should be 2:1 for not degraded RNA) as well as a virtual gel like image were obtained. Based on the quality measurements degradation of total RNA like in the formaldehyde gel-electrophoresis could be detected. Clear signs for degradation were decreasing ratio of ribosomal bands, additional peaks below the ribosomal bands, decrease in overall RNA signal intensity and a shift towards shorter fragments. All these facts are combined in the RNA integrity number (RIN) which can be computed by the software. The RIN takes into account the shape of the electropherogramm and assigns integrity quantitative values to the RNA measurements. The algorithm used allows a user-independent, automated and reliable analysis for a standardized RNA quality control (Imbeaud *et al.*, 2005; Schroeder *et al.*, 2006). RNA samples are sorted into ten different RNA integrity categories ranging from 1 (worst) to 10 (best). A subsequent experiment like microarray or qRT-PCR was only performed with samples having a RIN higher then 7.5. Cell culture samples mostly had values between 8 and 9 while samples from frozen human liver samples often showed a higher degradation state, sometimes even lower than 7.5.

2.8.4 Precipitation of Nucleic Acids

solutions: 3.0 M sodium acetate (NaOAc): $C_2H_3NaO_2$ $M_r = 82.04$

$$n = \frac{m}{M} \quad m = n \cdot M = 3.0 \cdot 82.04 = \underline{\underline{246.12 \text{ g}}} \text{ for 1 l}$$

$NaC_2H_3O_2$: 246.1 g 2.46 g

ddH₂O: 800.0 ml 8.00 ml adjust pH to 5.2

ad ddH₂O: 1000.0 ml 10.00 ml

3 M sodium acetate (pH 5.2): 408.1 g of sodium acetate x 3 H₂O in 800 ml Sodium acetate solved into water and pH to 5.2 adjusted using glacial acetic acid. The volume was filled to one litre with water. The Solution was autoclaved to sterilize it and frozen in aliquots.

Ethanol 100% und 75%

To obtain 5 - 25 µg total RNA (depending on the protocol) without amplification the RNA was precipitated by ethanol with high salt concentration. Precipitation was done using 1/10 volume of 3 M NaOAc, pH 5.2 (2.46 g ad. 10 ml) and 2.5 volume 100 % ethanol. The mixture was incubated for at least 1 h at -20°C. During the subsequent step the RNA was pelletized at 20,000 g ($\geq 12,000$ g) and 4°C for 20 min. The pellet was washed two times (mixed and pelletized) using 500 µl 70 % ethanol and then dried in the vacuum centrifuge at room temperature. Measuring the loss of weight the drying process could be observed to prevent over drying. The dried pellet was resolved in an appropriate amount of RNase free water.

2.8.5 RNA Amplification

The RNA amplification was done using the MessageAmp™ II aRNA Kit von Ambion which is based on an RNA amplification protocol developed in the laboratory of Dr. James Eberwine (Van Gelder *et al.*, 1990). An overview of the experimental procedure is given in Figure 25.

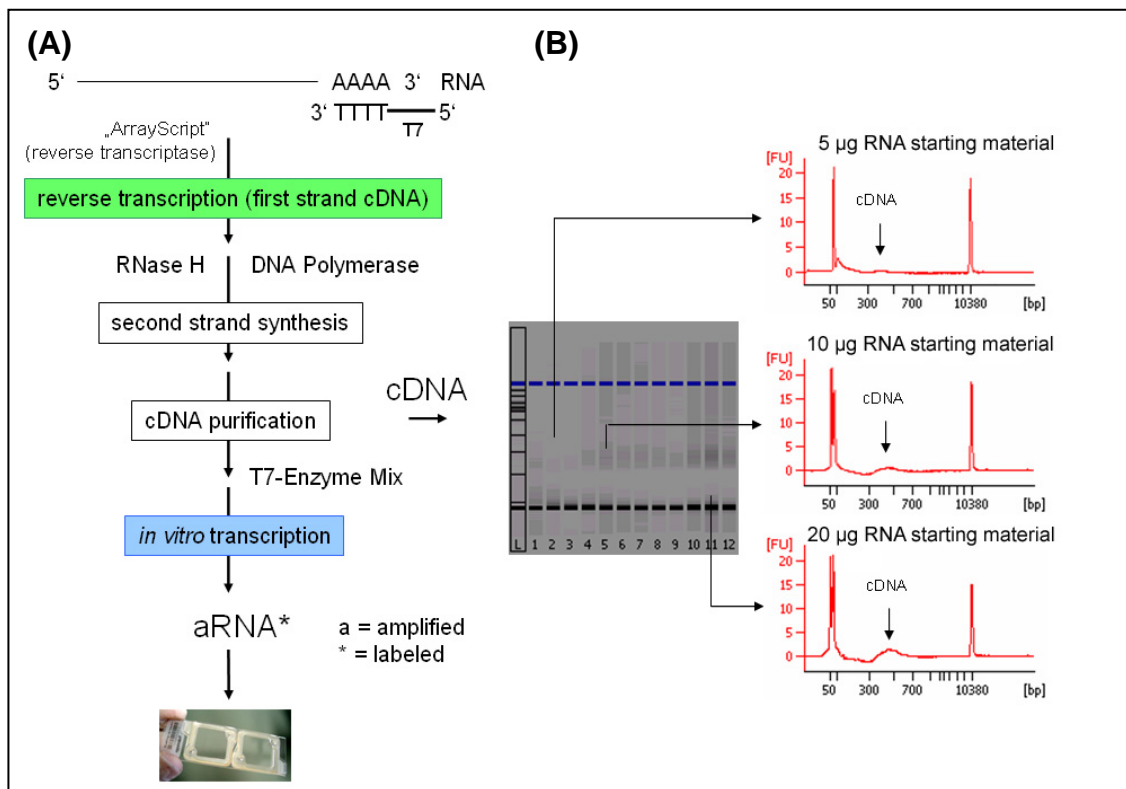


Figure 25: MessageAmp II RNA amplification procedure based on the "Eberwine protocol" (Van Gelder *et al.*, 1990) and analysed cDNA using the Agilent Bioanalyzer. (A) The amplification procedure consists of reverse transcription with an oligo(dT) primer bearing a T7 promotor and *in vitro* transcription of the resulting cDNA with T7 RNA polymerase to generate several thousand anti sense RNA copies of each RNA sample. The aRNA is also labeled during this procedure by using biotin labeled nucleotides. The labeled and purified aRNA could be directly hybridized on the microarray. As too less cDNA was obtained after reverse transcription (B) the generated cDNA could not be used directly for hybridization but had to be amplified first, using *in vitro* transcription.

The procedure consists of reverse transcription with an oligo(dT) primer bearing a T7 promotor and *in vitro* transcription (IVT) of the resulting cDNA with T7 RNA polymerase to generate hundreds to thousands of anti sense RNA copies of each RNA sample (in this work the anti sense RNA is referred to as aRNA, in the literature it is commonly called cRNA). RNA amplification has become the standard method for preparing RNA samples for array analysis (Kacharina *et al.*, 1999; Pabon *et al.*, 2001).

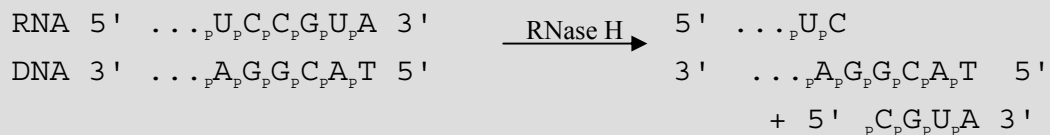
The first strand cDNA synthesis reaction employs the patent pending ArrayScript™ reverse transcriptase. The basic principle of such an enzyme is described under 2.8.9. During the second strand synthesis the RNA of the RNA DNA hybrid is digested. The activity of the RNase H is summarized in the following overview:

RNase H

Activity: 5' → 3' and 3' → 5' exoribonuclease

Substrate: degrades RNA in an RNA:DNA hybrid by processive mechanism

Reaction: Degradation of substrates with free ends yielding ribonucleotide products that are 4 - 20 nucleotides in length and contain 5'-phosphate and 3'-hydroxyl termini.



from Sambrook et al., 1989

For preparation of the master mixes the master mix calculator on the Ambion homepage (www.ambion.com/techlib/append/msgamp2_mm_calc.php) which calculates the appropriate amounts needed for your reactions was used. As an example the master mix for 4 reactions using one round of amplification and biotin labeling in a 40 µl reaction are given in the following tables (Table 12 - Table 15):

Table 12: Reverse Transcription Master Mix for RNA Amplification with MessageAmp™ II aRNA Kit from Ambion (amounts are given for four samples with 5% excess)

Reverse Transcription Master Mix		
Assemble at room temperature, adding reagents in the order shown		
Component	per reaction	Reagent Amounts (µl)
10x First Strand Buffer	2	8.4
dNTP Mix	4	16.8
RNase Inhibitor	1	4.2
ArrayScript	1	4.2
Total (µl)	8	33.6
Use 8 µl of Reverse Transcription Master Mix per reaction.		

Table 13: Second Strand Master Mix for RNA Amplification with MessageAmp™ II aRNA Kit from Ambion (amounts are given for four samples with 5% excess)

Second Strand Master Mix		
Assemble on ice, adding reagents in the order shown		
Component	per reaction	Reagent Amounts (µl)
Nuclease-free Water	63	264.6
10x Second Strand Buffer	10	42
dNTP Mix	4	16.8
DNA Polymerase	2	8.4
RNase H	1	4.2
Total (µl)	80	336
Use 80 µl of Second Strand Master Mix per reaction.		

Table 14: IVT Master Mix for RNA Amplification with MessageAmp™ II aRNA Kit (amounts are given for four samples with 5% excess)

IVT Master Mix for 40 µl reaction		
Assemble at room temp, adding reagents in the order shown		
Component	per reaction	Reagent Amounts (µl)
T7 ATP Solution	4	16.8
T7 CTP Solution	4	16.8
T7 GTP Solution	4	16.8
T7 UTP Solution	3	12.6
T7 10x Reaction Buffer	4	16.8
T7 Enzyme Mix	4	16.8
Total (µl)	22	92.4
Use 22 µl of IVT Master Mix per reaction.		

Usually the economic protocol (20 µl) of the IVT Master Mix was used. Preparing a master mix for four reactions with 5% excess the following amounts were needed:

Table 15: Economic protocol of the IVT Master Mix for RNA Amplification with MessageAmp™ II aRNA Kit von Ambion

IVT Master Mix for 20 µl reaction (economic protocol)		
Assemble at room temp, adding reagents in the order shown		
Component	per reaction	Reagent Amounts (µl)
T7 ATP Solution	2	8.4
T7 CTP Solution	1.5	6.3
T7 GTP Solution	2	8.4
T7 UTP Solution	1.5	6.3
T7 10x Reaction Buffer	2	8.4
T7 Enzyme Mix	2	8.4
Total (µl)	11	46.2
Use 11 µl of IVT Master Mix per reaction.		

To generate amplified RNA 1 µg total RNA was given into a sterile, RNase-free 0.5 ml microcentrifuge tube and 1 µl of T7 Oligo (dT) primer was added. The total volume was filled with nuclease-free water to 12 µl. This was mixed and then centrifuged to collect the mixture at the bottom of the tube.

The well mixed primer template solution was incubated for 10 min at 70°C to resolve secondary structures of the RNA and afterwards placed on ice.

Meanwhile the Reverse Transcription Master Mix was prepared at room temperature in a nuclease-free tube according to Table 12. The mixture was mixed well and briefly centrifuged to collect the mix on the bottom of the tube.

8 µl of Reverse Transcription Master Mix (Table 12) were transferred to the RNA sample primer mixture and the reaction was thoroughly pipetted up and down three times and in addition the tube was flicked around three to four times, followed by a quick spin to collect the reaction solution at the bottom of the tube. Finally the reaction mixture was placed at 42°C for 2 hours. At the end of the incubation the Second Strand Master Mix (Table 13) was prepared on ice, mixed and centrifuged.

After two hours of incubation the samples were spun briefly to collect the reaction mix at the bottom of the tube and placed on ice. The temperature of the incubator was lowered to 16°C. The workflow continued immediately by the addition of 80 µl Second Strand Master Mix to each sample. After addition of master mix the tubes were incubated at 16°C (lid temperature 20°C) for 2 hours in a thermal cycler. Nuclease-free water was preheated to 54°C. After 2 hrs incubation at 16°C, the reactions were placed on ice. For purification of cDNA 250 µl of cDNA Binding Buffer were added to each sample from the previous step and mixed thoroughly by pipetting up and down 2 - 3 times and flicking the tube 3 - 4 times, followed by a quick spin to collect the reaction mix at the bottom of the tube. The cDNA sample / cDNA Binding Buffer was transferred onto the centre of the cDNA filter cartridge and centrifuged for 1 min at 10,000 x g. The flow-through was discarded and the cDNA filter cartridge was replaced in the wash tube. 500 µl cDNA Wash Buffer were applied to each cDNA filter cartridge. The cartridge was centrifuged for 1 min at 10,000 x g until all the cDNA Wash Buffer is through the filter. The flow-through was discarded and the cartridge was spun for an additional minute to remove trace amounts of ethanol. The cDNA Filter Cartridge was finally transferred to a cDNA elution tube.

To elute the cDNA 10 μ l of pre-heated nuclease-free water (54°C) was applied to the centre of the filter in the cDNA Filter Cartridge which was further incubated for 2 min and then centrifuged for 1.5 min at 10,000 x g. The elution was repeated with a second 10 μ l of preheated nuclease-free water. Finally the double stranded cDNA was found in the eluate (~ 16 μ l).

Before starting with the in vitro transcription the cDNA was mixed with the biotinylated nucleotides and concentrated. 7.5 μ l of biotin labeled NTPs were added to the eluted cDNA, and the mixture was concentrated in a vacuum centrifuge concentrator until the volume was reduced to 9 μ l for 20 μ l reaction. This took approximately 10 minutes at 30°C. To avoid over concentration or even drying the volume reduction was followed by measuring the loss of weight in a balance (Table 16).

Table 16: Loss of weight during concentration of cDNA biotin labeling mix

time [min]	sample 1 [g]	Δ volume [g]	sample 2 [g]	Δ volume [g]
	1.042		1.057	
6	1.036	0.006	1.051	0.006
7	1.029	0.013	1.044	0.013
9	1.021	0.021	1.036	0.021

10 μ l = 0.01 g; reduced by approximately 20 μ l (remaining V = ~ 10 μ l)

At room temperature, the prepared IVT Master Mix was added to each sample. The reaction was mixed thoroughly by pipetting up and down 2-3 times, then flicking the tube 3-4 times, followed by a quick spin to collect the reaction mixture in the bottom of the tube. Once assembled, the tubes were placed in a 37°C incubator for 5 hours. The minimum recommended incubation time is 4 hours and the maximum is 14 hours. It was very important to maintain a constant 37°C incubation temperature. Therefore a PCR cycler was used for this step with a heated lid at 39°C.

The reaction was stopped by adding nuclease-free water to each aRNA sample to bring the final volume to 100 μ l. Before beginning the aRNA purification, the nuclease-free water was again heated to 54°C for at least 10 min. For each sample, an aRNA Filter Cartridge was placed into an aRNA collection tube, and set aside. 350 μ l aRNA Binding Buffer were added to each sample followed by the addition of 250 μ l 100 % ethanol. The solution was mixed by pipetting up and down 3 times. At this point it was

important to avoid any vigorous mixing and centrifugation. Immediately after addition of ethanol and slight mixing the samples were applied to the prepared aRNA Filter Cartridges. Any delay in proceeding could result in loss of aRNA because once the ethanol was added the aRNA was in a semi-precipitated state.

The cartridges with the samples were centrifuged for 1 min at 10,000 x g. The flow-through was discarded and the aRNA Filter Cartridge replaced back into the aRNA collection tube. 650 µl Wash Buffer were applied to each aRNA Filter Cartridge and centrifuged for min at 10,000 x g. The flow-through was discarded and the cartridge was centrifuged for an additional 1 min to remove trace amounts of ethanol. The Filter Cartridges were transferred to a fresh aRNA collection tube.

To elute the purified aRNA 100 µl nuclease-free water, preheated to 54°C, were added to the centre of the filter. The device was incubated at room temperature for 2 min and then centrifuged for 1.5 min at 10,000 x g. The aRNA was then found in the aRNA collection tube in approximately 100 µl of nuclease-free water.

The concentration and quality was checked by measurement of absorbance and gel-electrophoresis (see 2.8.3).

The aRNA samples were stored at -80°C for up to 1 year. To minimize repeated freeze-thawing the samples were split into 5 – 20 µg aliquots.

2.8.6 Digestion of DNA

The digestion of DNA was done using the recommended conditions of the supplier of the enzyme. An example is given in the following pipetting scheme of a digestion reaction:

plasmid DNA (miniprep):	4 µl (2 µg)
10 x buffer (enzyme specific here "L"):	2 µl
restriction enzyme e.g. <i>KpnI</i> [10 U/µl]:	1 µl
BSA (1 mg/ml sometimes already in the buffer):	2 µl
ddH ₂ O:	11 µl
<hr/>	
	20 µl

This set was incubated for 1.5 h at 37°C. To stop the digestion the mixture was heated to 65°C for 15 minutes. Adjacent the results were analysed using agarose gel electrophoresis (see 2.8.7).

2.8.7 Separation of Nucleic Acids Using Agarose Gel Electrophoresis

To verify degradation state, size and complete digestion of nucleic acid, agarose gel electrophoresis was used. First of all enough electrophoresis buffer (TAE) had to be prepared to fill the electrophoresis tank and prepare the gel. Usually 400 ml agarose was prepared at once. For a 1 % solution 4 g powdered agarose was added to 400 ml TAE buffer. In a glass bottle with a loose-fitting cap the slurry was heated in a microwave oven until the agarose dissolves. The dissolved agarose was stored at 70°C to keep it liquid. Otherwise it had to be resolved in the microwave again. About 40 ml of 1% agarose gel solution with 3 µl ethidium bromide were poured in a gel supporting tray to polymerize. A comb was applied to the tray which allowed forming of a well. After the gel was completely set (30 minutes at room temperature) it was mounted in the electrophoresis tank. Electrophoresis buffer (TAE) was added to cover the gel to a depth of about 1 mm. The sample comb was removed. The samples were mixed with sample buffer (1:5 with a 5-fold buffer). The mixture was slowly loaded into the slots of the submerged. The minimum of DNA that can be detected by photography of ethidium bromide stained gels is about 2 ng in a 0.5 cm wide band. If there is more than 500 ng of DNA in a band of this width, the slot will be overloaded, resulting in tailing and smearing. The lid of the gel tank was closed attaching the electrical leads. The DNA will migrate to the anode (red lead "+"). A voltage of $U = 80 \text{ V}$ (1 - 5 V/cm measured as the distance between the electrodes) was applied. During the start bubbles should be generated at the anode and cathode (due to electrolysis) and, within a few minutes, the bromophenol blue should migrate from the wells into the body of the gel. During electrophoresis, the ethidium bromide migrates toward the cathode (in the direction opposite to that of the DNA). Extended electrophoresis can remove much of the ethidium bromide from the gel, making detection of small fragments difficult. If the bands couldn't be properly observed, the gel was re-stained by soaking it for 30 minutes in a solution of ethidium bromide (0.5 µg/ml).

After a running time of 45 minutes (30 min to 1 h) the electric current was turned off. The gel was taken out of the tray, examined by ultraviolet light and photographed.

To separate RNA in an agarose gel formaldehyde agarose (FA) gel electrophoresis was performed. The basic protocol is based on Sambrook *et al.*, 1989 with some slight modifications (e.g. a more concentrated RNA loading buffer allowing a larger volume of RNA sample to be loaded onto the gel). The required solutions are given here:

buffer:	10x FA gel buffer:	200 mM MOPS (free acid)	10.50 g
	(250 ml)	50 mM sodium acetate	1.03 g
		10 mM EDTA (0.5 M SS*)	5.00 ml

adjust pH to 7.0 with NaOH

* 0.5 M EDTA-stock solution: 146.2 g / l is equivalent to 7.3 g / 50 ml

5x FA gel running buffer: 10x FA gel buffer dilute 1:2 with H₂O_{bidest}
add 0.1 % DEPC stir for 1 h and
autoclave (50 ml 10x FA gel buffer
+ 50 ml ddH₂O + 0.1 ml DEPC)

1x FA gel running buffer: 100 ml 5x FA gel buffer
10 ml 37 % (12.3 M) formaldehyde
440 ml DEPC treated ddH₂O

100 ml 10x FA gel buffer
20 ml 37 % (12.3 M) formaldehyde
880 ml DEPC ddH₂O

5x RNA loading buffer: 16 μ l saturated bromophenol blue solution¹
80 μ l 500 mM EDTA pH 8.0
720 μ l 37 % (12.3 M) formaldehyde²
2 ml 100 % glycerol
3084 μ l formamide
4 ml 10x FA gel buffer
ad. 10 ml DEPC treated ddH₂O

Stability: approximately 3 months at 4°C

¹ To make a saturated solution, add solid bromophenol blue to distilled water. Mix and continue to add more until no more will dissolve. To pellet the not dissolved powder centrifuge and carefully pipette the saturated supernatant.

² Toxic and/or mutagenic: take appropriate safety measures.

To prepare a FA gel (1.2 % agarose) 1.2 g agarose and 10 ml 10x FA gel buffer were combined and the volume was adjusted to 100 ml with nuclease-free water. The mixture was heated in the microwave to dissolve the agarose. After it had cooled down to approximately 65°C 1.8 ml of 37 % (12.3 M) formaldehyde and 1 μ l of a 10 mg/ml ethidium bromide stock solution (or 3 μ l of a 1 % in water solution) were added. The solution was thoroughly mixed and poured onto a gel support to polymerize. A comb was added to build wells for the samples. After solidification of the gel the comb was removed and prior to running the gel was equilibrated in 1x FA gel running buffer for at least 20 minutes. One volume of 5x loading buffer per 4 volumes of RNA sample (e.g. 10 ml of loading buffer and 40 μ l of RNA) were added and mixed. The mixture was incubated for 3 to 5 minutes at 65°C, to remove secondary structures, chilled on ice and loaded onto the equilibrated FA gel. The gel was run at 80 V (5 - 7 V/cm) in 1x FA gel running buffer. For comparison of size an RNA ladder was used (High Range 200 - 6000 bases, from Fermentas). The separated RNA could be detected by ultraviolet light and was photographed under UV light in.

2.8.8 Preparation and Transformation of Competent *E. coli*

The following simple procedure is a variation of that of Cohen *et al.*, 1972 and is frequently used to prepare batches of competent bacteria. The procedure, which works well with most strains of *E. coli*, is quick and reproducible. Competent cells made by this procedure were also preserved at - 80°C, although there may be some deterioration in the efficiency of transformation during prolonged storage.

Solutions:

preparatory culture 2 ml / glycerolstock *E. coli* DH5 α

LB-medium (1 L):	10 g Trypton	agar-plates: ad 15 g agar/litre
	5 g yeast extract	Ampicillin: 100 μ g/ μ l final concentration
	10 g NaCl	100 mg/ml stock solution

TfbI (200 ml):	component	amount	molar final concentration
	potassium acetate	0.59 g	30 mM
	rubidium chloride	2.42 g	100 mM
	calcium chloride (2 H ₂ O)	0.29 g	10 mM
	manganese chloride (4 H ₂ O)	2.00 g	50 mM
	glycerol (99.5 %)	30.00 ml	15 % v/v

adjust pH to 5.8 with diluted acetic acid (from about pH ~ 7) and sterilize by filtration*

TfbII (100 ml):	component	amount	molar final concentration
	MOPS	0.21 g	10 mM
	calcium chloride (2 H ₂ O)	1.10 g	75 mM
	rubidium chloride	0.12 g	10 mM
	glycerol (87 %)	17 ml	15 % v/v

adjust pH to 6.5 with diluted NaOH and sterilize by filtration or autoclave

* some components might precipitate during autoclaving

A single colony (2 - 3 mm in diameter) was picked from an agar plate freshly grown for 16 - 20 hours at 37°C and transferred into 2 to 5 ml LB-medium in a preparatory culture

tube. The tube was incubated over night at 37°C and 200 cycles / minute in a rotary shaker. After growth of about 16 hours 100 ml LB medium in a 1 litre flask was inoculated with 1 ml of the dense grown pre-culture ($OD_{600} \sim 10$; 1:100 \Rightarrow $OD_{600} \sim 0.1$).

The cells were again incubated at 37°C at 200 cycles / minute for about 2 hours. The growth of the culture was monitored by determination of OD_{600} about every 20 to 30 minutes (e.g. 8 a.m. $OD_{600} = 0.1$, 10 a.m. OD_{600} (1:2 diluted) = 0.3 \Rightarrow $OD_{600} = 0.6$). At OD_{600} of about 0.5 - 0.8 the culture was aseptically transferred to a sterile disposable, ice cold 50 ml polypropylene tube and cooled to 0°C by storing the tubes on ice for 10 to 15 minutes. All subsequent steps in this procedure were carried out aseptically. The cells were recovered by centrifugation at 3000 g and 4°C for 5 minutes (in 2 x 50 ml polypropylene tubes "Falcon"). The media was decanted from the cell pellets. The tubes were situated inverted for 1 minute to allow the last traces of media to drain away. Each pellet was resuspended in 0.4 volumes (original volume) TfbI (100 ml in 40 ml; 2 x 50 ml each in 20 ml) and incubated for 15 minutes in ice.

The cells were recovered by centrifugation at 3000 g at 4°C for 5 minutes. The supernatant was removed and traces of fluid drained away by inverting the tubes for 1 minute. The cell pellet was resuspended in 0.04 volumes of ice cold TfbII (2 times 2 ml). The suspension was incubated 15 minutes on ice. At this point the cells could be dispensed into aliquots by using a chilled, sterile pipette tip and transferring 50 μ l of the suspension of competent cells to a chilled, sterile microcentrifuge tube. Immediately after pipetting the competent cells were snap-frozen by immersing the tightly closed tubes in liquid nitrogen. The tubes were stored at - 80°C.

For transformation the competent cells were removed from the - 80°C freezer and thawed on ice. After the cells started to thaw, they were transferred to an ice bath and incubated for 10 minutes at 0°C. DNA (no more than 50 ng in a volume of 10 μ l or less e.g. 10 μ l ligation or 1 μ l plasmid DNA) was added to each tube. With every transformation there were always two control experiments performed. On one hand competent bacteria that receive a known amount of a standard preparation of circular plasmid DNA (as positive control) and on the other hand competent bacteria that receive no plasmid DNA at all (water - as negative control). The contents of the tubes were gently mixed by swirling and stored on ice for 30 minutes. After that incubation

step the tubes were transferred in a heating block or water bath which was pre- heated to 42°C. The tubes were incubated for exactly 45 sec. Following this heat shock the cells were immediately transferred to an ice bath avoiding vigorous shaking. The cells were chilled for 1 minute. After this short cooling step 500 µl of LB medium was added to the cells and they were incubated for 1 h at 37°C to allow the bacteria to recover and to express the antibiotic resistance marker encoded by the plasmid. The appropriate volume (up to 200 µl per 90 mm plate) of transformed competent cells was transferred on selective agar plates with the dedicated antibiotic. A sterile bent glass rod was used to gently spread the transformed cells over the surface of the agar plate. If more than 50 µl of transformed cells had to be plated on a single 90 mm plate the cells were concentrated by centrifugation and resuspended in a smaller amount (5 - 20 µl) of LB media. However, in the case of resistance to Ampicillin, only a portion of the culture (empirically determined - several amounts e.g. 30 µl and 10 µl) was spread on a single plate, (at least for transformation of a circular plasmid) avoiding to get too many colonies.

The plates were left at room temperature until the liquid was absorbed and then they were inverted and incubated at 37°C. Usually colonies appeared in 12 - 16 hours. For resistance to Ampicillin it had to be considered that cells should be plated at low densities ($<10^4$ colonies per plate) and the plates should not be incubated for more than 20 hours at 37°C. β -lactamase, secreted into the medium from Ampicillin resistant transformants, rapidly inactivates the antibiotic in regions surrounding the colonies. Thus, plating cells at high density or incubating them for long periods results in the appearance of Ampicillin sensitive satellite colonies.

2.8.9 Reverse Transcription

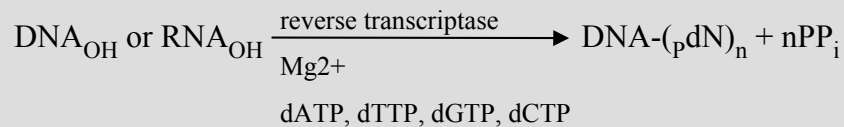
Reverse transcriptase is used to transcribe mRNA into double stranded cDNA which was used in hybridization experiments. The use of labeled nucleotides allows the labeling of cDNA during this procedure. An overview of reverse transcriptase activity and reaction principle is given in the following scheme:

Reverse Transcriptase

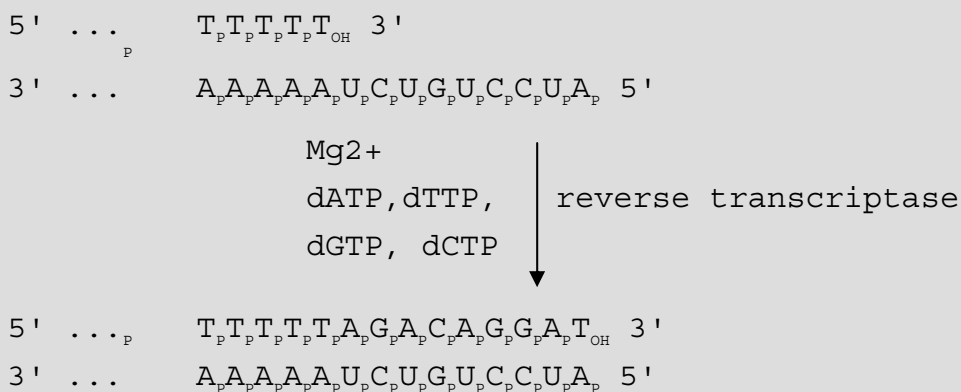
Activity: 5' → 3' DNA polymerase

Substrate: RNA or DNA template with a RNA or DNA primer bearing a 3'-hydroxyl group.

Reaction:



For example:



from Sambrook et al., 1989

The purified RNA was reversely transcribed into cDNA using the reverse transcriptase "Superscript III RNase H reverse transcriptase" from Invitrogen. Although there are three types of primers (oligo dT, oligonucleotides of random sequence Taylor *et al.*, 1976 and oligonucleotides of defined sequence) possible to use, in this study only oligo dT primers were used.

The following components were added to a nuclease-free microcentrifuge tube:

	1 µl	500 ng Oligo (dT)12-18 Primer (0.5 µg/µl)
	1 µl	10 mM dNTP Mix (10 mM each dATP, dCTP, dGTP, dTTP)*
	x µl	total RNA (1 µg)
ad.	13 µl	sterile, distilled water

* (2.5 µl dATP, dCTP, dGTP, dTTP = 10 µl + 90 µl H₂O 100 µl 10 mM dNTP)

For a labeling reaction, where the generated cDNA should be labeled with a fluorescent dye or biotin the dNTP mix was adapted (see 2.10.1).

The mixture was heated to 65°C for 5 minutes and then incubated on ice for at least 1 min. The contents of the tube were collected by brief centrifugation. The following components were added to the RNA primer mixture:

- 4 µl 5x First-Strand Buffer
- 1 µl 0.1 M DTT
- 1 µl rRNasin RNase Inhibitor (or RNase_{OUT}TM)
- 1 µl SuperScriptTM III RT (200 U/µl)

The solution was mixed by pipetting gently up and down and the tube was incubated at 25°C for 5 minutes. Subsequently the reaction was incubated at 50°C for 45 (30-60) minutes. To stop the reaction it was inactivated by heating to 70°C for 15 minutes. The cDNA generated could be stored at -20°C.

2.8.10 Quantitative Real-time PCR

Using a quantitative real-time PCR assay it is possible to amplify and detect a PCR product simultaneously in real time. Such a format is e.g. the so called TaqManTM assay (Applied Biosystems, Darmstadt). In this assay the 5' - 3' exonuclease activity of the Taq DNA polymerase (Holland *et al.*, 1991; Lyamichev *et al.*, 1993) is exploited to release a fluorescence dye which was hybridized as TaqMan probe between the two primers and which is cleaved during the amplification by the polymerase. A schematic example of such a probe, which was developed by Applied Biosystems (Lee *et al.*, 1993) is shown in Figure 26.

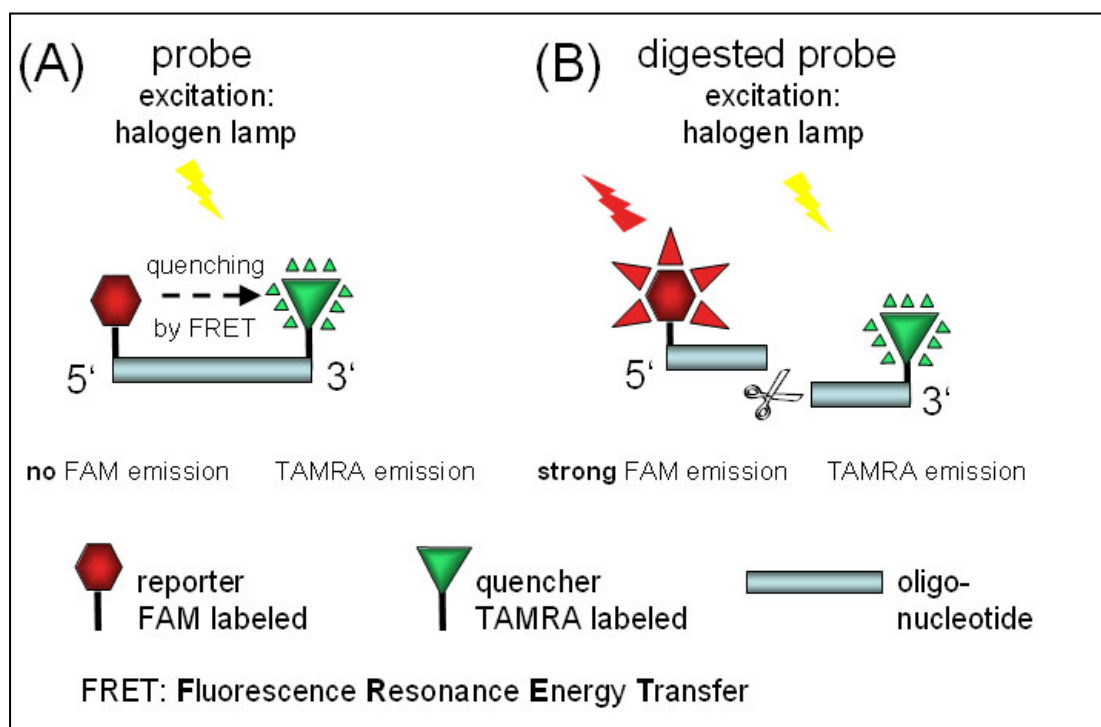


Figure 26: Basic principle of a TaqMan probe. The fluorogenic probe consists of an oligonucleotide (blue bar) with a reporter fluorescent dye (here FAM = 6-Carboxy-Fluorescein, emission 518 nm) attached to the 5' end and a quencher dye (here TAMRA = 6-Carboxy-tetramethyl-rhodamine, emission 582 nm) at the 3' end. (A) When the FAM is excited by irradiation, its fluorescence emission will be quenched if the TAMRA is close enough to be excited through the process of fluorescence energy transfer (FRET). (B) During PCR, if the probe is hybridized to a template strand, Taq DNA polymerase will cleave the probe. If cleavage occurs between the FAM and TAMRA dyes it causes an increase in FAM fluorescence intensity because the FAM is no longer quenched.

If such a probe (Figure 26) is hybridized to a template strand, the DNA Taq polymerase will cleave during the amplification step of the PCR the probe due to its inherent 5' - 3' nucleolytic activity. Consequently the release of the marker dye from the quencher causes an increase in fluorescence intensity because the quenching is completely dependent on the physical proximity of the two dyes (Stryer and Haugland, 1967). The increase in reporter fluorescence intensity indicates that the probe specific PCR product has been generated. This amplification and detection step is diagrammed in Figure 27.

To validate the array results the quantitative determination of CYP3A4, CYP2B6 and 18S rRNA expression was analysed by real-time PCR using ABI Prism™ 7500 Sequence Detector (Applied Biosystems) and Mastercycler Gradient Realplex (Eppendorf).

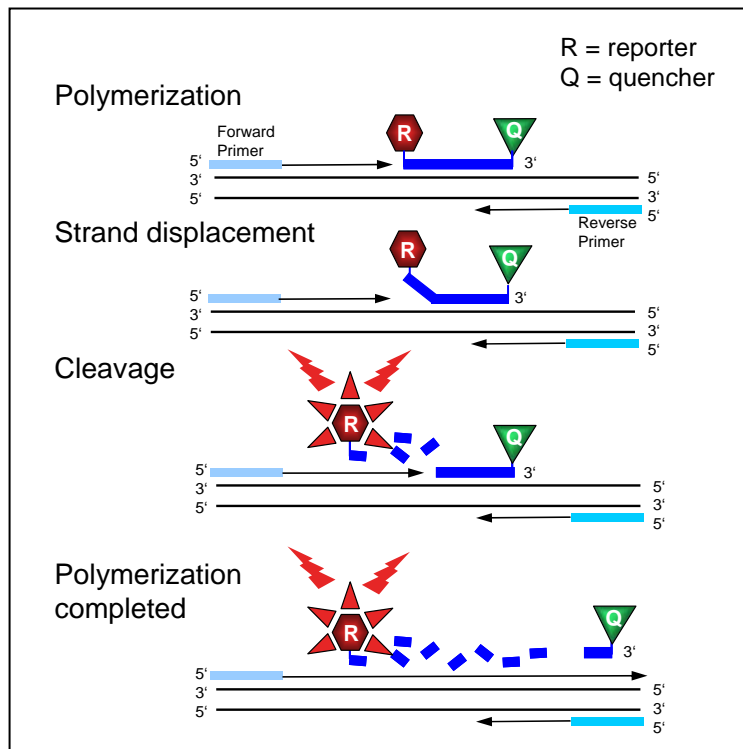


Figure 27: Diagram of TaqMan principle. Step wise representation of the 5' - 3' nucleolytic activity of Taq DNA polymerase acting on a fluorogenic probe during one extension phase of real-time PCR step cycle.

The primer- and probe sequences used for the CYP3A4 were the following:

CYP3A4

forward primer 5'-TGT CCT ACC ATA AGG GCT TTT GTA T-3'

reverse primer 5'-TTC ACT AGC ACT GTT TTG ATC ATG TC-3'

probe 5'-CTT TTA TGA TGG TCA ACA GCC TGT GCT G-3'

In addition to the customized assay for CYP3A4 and 18S rRNA pre-developed assay reagent Kits (PDAR) from Applied Biosystem were used. The PCR was done in 96-well plates (Applied Biosystems, ABgene and Eppendorf) in a final volume of 25 μ l. First of all the mRNA sample had to be transcribed into cDNA. This was done using 19.25 μ l of mRNA (1 μ g) in water, adding 30.75 μ l of RT Master Mix (Table 17) and incubation at the following temperature profile: 10 min 25°C, 30 min 48°C, 5 min 95°C. The cDNA samples were portioned and could be stored at -20°C. For real-time PCR the cDNA samples had to be diluted. For the CYP3A4 and CYP2B6 assay the

dilution was 1:25 (3 μ l cDNA and 72 μ l water) and for the 18S rRNA assay the dilution was 1:40 (3 μ l of a 1:25 dilution plus 117 μ l water). To 5 μ l of diluted cDNA template 20 μ l TaqMan Master Mix containing 12.5 μ l 2 \times Universal PCR Master Mix (Applied Biosystems), 5 μ l cDNA, reverse and forward primer (Applied Biosystems and Metabion) as well as the probe (Applied Biosystems) were added. The RNA primer- and probe-concentrations used for a real-time PCR experiment can be found in Table 18. The reaction was performed using the temperature profile given in Figure 28. The first step for 2 minutes at 50°C is the so called "UNG"-step. The primer/probe/nucleotide mix contains dUTP instead of dTTP. Using the heat unstable enzyme uracil-DNA glycosylase (UNG) prevents a contamination of the PCR reaction from previous runs. In all runs dTTP was exchanged by dUTP. Therefore undesired PCR templates were removed by cleavage at the UTPs and subsequently hydrolyzed during denaturation. In this denaturation step the enzyme (UNG) will be inactivated. Native DNA is not affected, because it has no uracil and serves as template in the PCR reaction.

Table 17: Composition of RT Master Mix for cDNA synthesis reaction from 200 ng RNA for real-time PCR.

component	c(stock)	volume	c(final)
TaqMan RT buffer	10x	5.00 μ l	1x
magnesium chloride	25 mM	11.00 μ l	5.5 mM
desoxyNTP-Mix	2.5 mM	10.00 μ l	500 μ M each
primer (hexamer)	50 μ M	2.50 μ l	2.5 μ M
RNase inhibitor	20 U/ μ l	1.00 μ l	0.4 U/ μ l
Multiscribe RT	50 U/ μ l	1.25 μ l	1.25 U/ μ l
RNA		x μ l	200 ng
RNase free water		3.85 - x μ l	

Table 18: Used oligo nucleotide concentration and amount of RNA for qRT-PCR.

	forward primer c _{final} [nM]	reverse primer c _{final} [nM]	probe c _{final} [nM]	dye of probe	amount of used RNA
CYP3A4/2B6	400	400	200	FAM	40 ng
18S rRNA	assay reagent Kit (Applied Biosystems)			VIC	40 pg

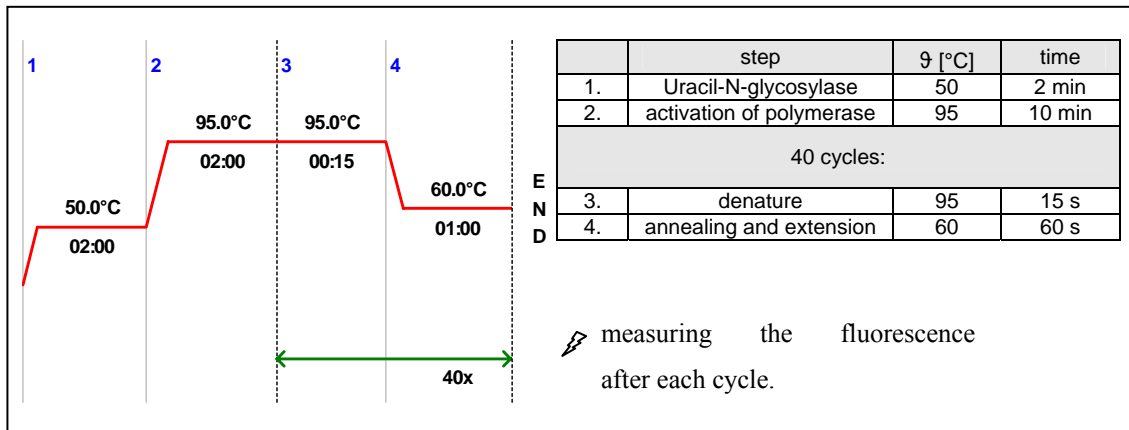


Figure 28: Temperature profile of real-time PCR (from the Mastercycler gradient realplex, Eppendorf). The left side displays on graphical overview of the PCR cycle program while the right side indicates detailed information about the temperature and time of the different PCR cycles.

2.8.10.1 Calculation of a CYP3A4 standard curve for qRT-PCR

The calibration curve was set up from 30 to 300 000 copies of the target gene CYP3A4 which was cloned in a vector giving the total construct a size of 6 kbp. The concentration of the DNA from the digested stock solution was determined using the Nanodrop photometer, yielding 115 ng/μl. The concentration was calculated to have the final copy number in a volume of 5 μl plasmid DNA / PCR reaction.

First of all the mass of the single plasmid molecule was determined:

$$m_{Plasmid} = n \cdot 1.096 \cdot 10^{21} \frac{g}{bp}$$

$$m_{Plasmid} = 6000 \text{ bp} \cdot 1.096 \cdot 10^{21} \frac{g}{bp} = \underline{\underline{6.58 \cdot 10^{-18} g}}$$

The amount of plasmid was calculated to achieve the desired copy number:

$$300\,000 \text{ copies equal to } 6.58 \cdot 10^{-18} \frac{g}{copy} \cdot 300,000 \text{ copies} = \underline{\underline{1.97 \cdot 10^{-12} g}}$$

The needed plasmid DNA (1.97 pg) was divided by the volume that is used per reaction to achieve the equivalent copy number (300,000):

$$1,97 \cdot 10^{-12} \text{ g} \div 5 \mu\text{l} = \underline{\underline{3,95 \cdot 10^{-12} \text{ g} / \mu\text{l}}}$$

Calculation of dilution series:

per dilution:	starting concentration:	115	ng/ μl
	1:100	1.15	ng/ μl
	1:100	11.5	pg/ μl
	1: 10	1.15	pg/ μl $1.15 \cdot 10^{-12} \text{ g} / \mu\text{l}$

$$c_1 \cdot V_1 = c_2 \cdot V_2$$

$$1.15 \cdot 10^{-12} \text{ g} / \mu\text{L} \cdot V_1 = 3.95 \cdot 10^{-13} \text{ g} / \mu\text{L} \cdot 100 \mu\text{L} \Rightarrow V_1 = \frac{3.95 \cdot 10^{-13} \text{ g} / \mu\text{L} \cdot 100 \mu\text{L}}{1.15 \cdot 10^{-12} \text{ g} / \mu\text{L}} = \underline{\underline{34.3 \mu\text{L}}}$$

The calculations are summarized in Table 19 showing the pipetting scheme for the dilution series of CYP3A4 plasmid calibration curve:

Table 19: Pipetting scheme for a CYP3A4 plasmid DNA calibration curve for quantitative real-time PCR

step	stock solution	concentration [g/ μl]	volume [μl]	volumen H ₂ O [μl]	final-volume [μl]	endconcentration [g/ μl]	copy number / 5 μl
	"Mini"						
1	(digest)	0.12×10^{-6}	10	990	1000	0.12×10^{-8}	N/A
2	dilution 1	0.12×10^{-8}	10	990	1000	0.12×10^{-10}	N/A
3	dilution 2	0.12×10^{-10}	10	90	100	1.20×10^{-12}	N/A
4	dilution 3	1.20×10^{-12}	34.3	65.7	100	3.95×10^{-13}	300000
5	dilution 4	3.95×10^{-13}	10	90	100	3.95×10^{-14}	30000
6	dilution 5	3.95×10^{-14}	10	90	100	3.95×10^{-15}	3000
7	dilution 6	3.95×10^{-15}	10	90	100	3.95×10^{-16}	300
8	dilution 7	3.95×10^{-16}	10	90	100	3.95×10^{-17}	30

2.9 Protein Biochemical Methods

2.9.1 Protein Precipitation with Acetone

As cell material was always a limiting factor a protocol was used to gain protein fractions from the same samples used for RNA extraction. This protocol was designed for acetone precipitation of proteins from cell lysates of primary human hepatocyte cells using buffer RLT (supplied with RNeasy Kits for RNA isolation). The precipitated, denatured protein was suitable for applications like SDS-PAGE (2.9.2) or western blotting (2.9.3).

The flow through from the RNA purification (protocol given at 2.8.2) was mixed with four volumes of ice-cold acetone (200 μ l flow through to 800 μ l acetone). The mixture was incubated 20 min on ice or -20°C . The protein was precipitated by centrifugation for 10 minutes at maximum speed in a benchtop centrifuge. The supernatant was discarded and the pellet was air-dried. The pellet was resuspended in 2 x SDS gel-loading buffer to give a final concentration of one fold. The samples could be frozen or immediately loaded on a SDS-PAGE after heating to 95°C for 3 minutes.

2.9.2 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is used to assess the purity and estimate the molecular weight of a protein. SDS-PAGE was done based on the protocol from Laemmli, 1970. The following solutions are required for SDS-PAGE analysis:

SDS gel electrophoresis buffers:

Tris-glycine	stock solution (per litre 5x)	working solution 1 x
	15.1 g Tris base (not Tris x Cl)	25 mM Tris 250 ml 5x stock
	94.0 g glycine (electrophoresis grade)	250 mM glycine 1000 ml ddH ₂ O
	50.0 ml 10 % SDS (electrophoresis grade)	0.1 % SDS

add ddH₂O to 1 litre, and adjust pH to 8.3.

2 x SDS gel-loading buffer	100 mM Tris HCl (pH 6.8)	1 M Tris 1 ml
	200 mM dithiothreitol ⁺	
	4.0 % SDS (electrophoresis grade)	400 mg SDS
	0.2 % bromophenol blue (bpb)	20 mg bromophenol blue
	20.0 % glycerol	2 g (2.3 g 87 %)
ad. 10.0 ml ddH ₂ O		

⁺ 2x SDS gel-loading buffer lacking dithiothreitol can be stored at room temperature. Dithiothreitol should then be added, just before the buffer is used, from a 1 M stock (200 mM: 50 µl of a 1 M stock solution DTT / 200 µl loading buffer).

1 M Dithiothreitol (DTT): Dissolve 3.09 g of DTT in 20 ml of 0.01 M sodium acetate (pH 5.2; M_r 82.04; 0.016 g/20 ml – 10x 0.16 g/20 ml). Sterilize by filtration (do not autoclave!). Dispense into 1 ml aliquots and store at -20°C.

10% Sodium dodecyl sulphate: Dissolve 100 g of electrophoresis grade SDS in 900 ml of H₂O; heat to 68°C to assist dissolution; adjust the pH to 7.2 by adding a few drops of concentrated HCl; adjust the volume to 1 litre with H₂O; dispense into aliquots. There is no need to sterilize 10 % SDS.

Ammonium persulphate: 10 % (w/v) to 1 g of ammonium persulphate, add H₂O to 10 ml. The solution may be stored for several weeks at 4°C.

1.0 M Tris: Dissolve 121.1 g of Tris base in 800 ml of H₂O. Adjust the pH to the desired value by adding concentrated HCl (12.11 g / 100 ml – 50 ml pH 8.8 and 50 ml pH 6.8):

pH	7.4	7.6	8.0
HCl	70 ml	60 ml	42 ml

Solution has to cool to room temperature for final pH adjustment! Adjust the volume to 1 litre with H₂O. Sterilize and dispense into aliquots.

First of all the glass plates including sealing were cleaned and assembled. The resolving gel solution was prepared, mixing the components in the order shown in Table 20. The polymerization was started by adding TEMED. Without delay the acrylamide mixture was swirled and rapidly poured into the gap between the glass plates. The gap was not

completely filled to leave sufficient space for the stacking gel (about 1.5 cm). The gel mixture was overlaid with isopropanol or 0.1 % SDS. The stacking gel solution was prepared adding the components from Table 21. After complete polymerization of the resolving gel the overlay was poured out and the stacking gel solution was directly put onto the surface of the polymerized resolving gel. A comb was inserted into the stacking gel solution to form the cavities for the samples. While the stacking gel was polymerizing, the samples were prepared by heating them to 95°C for 3 min in 1x SDS gel-loading buffer. After polymerization the wells were washed with water or buffer, the gel was mounted in the electrophoresis apparatus, and Tris-glycine buffer was added. The samples, 5 µl of a protein standard (mixture of coloured 9 proteins - PageRuler™ Prestained Protein Ladder Plus, Fermentas; Figure 29) were loaded and an equal volume of 1 x SDS gel loading buffer was put into any wells that were unused. The power was set to 8 V/cm for 10 min (10 mA/gel) for the initial phase and after 10 min it was set to 15V/cm for further 60 minutes (25 mA/gel ~ 50 V). If two gels were run in parallel the values were 10 min at 50 V, 20 mA, 1 W and the second phase 60 min 120 V, 50 mA, 6W.

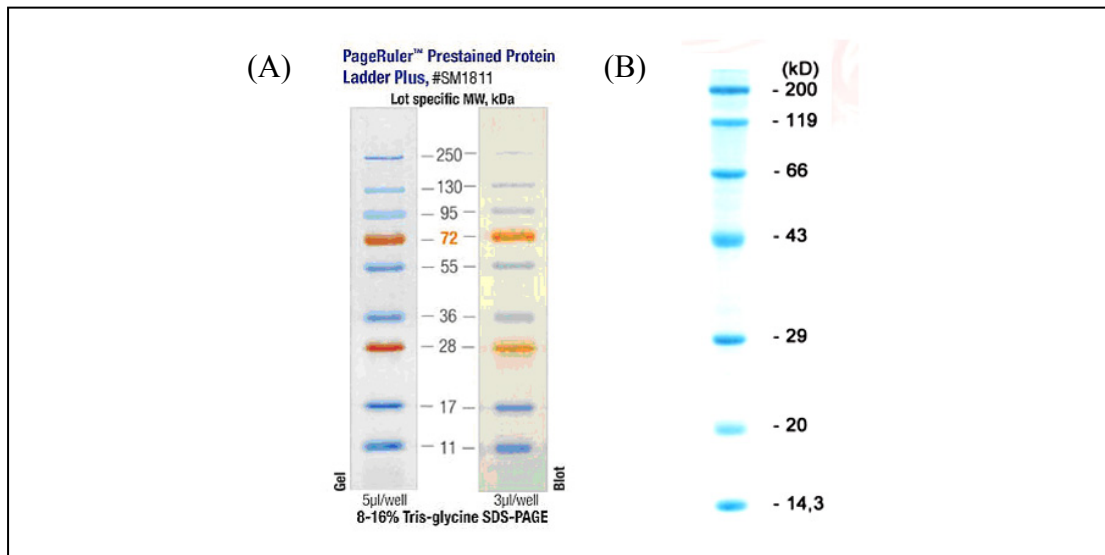


Figure 29: PageRuler™ Prestained Protein Ladder Plus (A) for monitoring protein separation during SDS-polyacrylamide gel electrophoresis, verification of Western blot transfer efficiency and approximate sizing of proteins. It is a mixture of 9 recombinant, highly purified, coloured proteins with apparent molecular weights of 10 to 250 kDa. A blue chromophore is coupled to these proteins, and proteins of two reference bands of ~70 kDa and of ~27 kDa are coloured with an orange dye. (B) Roti®-Mark STANDARD with a size distribution from 14.3 - 200 kDa with 7 protein fragments.

Table 20: Tris-glycine SDS-polyacrylamide gel electrophoresis: solutions for 12% resolving gel. Shown in grey are the normally used amounts.

Solution components	Component volumes [μ l] per gel volume of			
	5 ml	10 ml	15 ml	20 ml
12 %		1 gel	2 gels	
H ₂ O	1600	3300	4900	6600
30 % acrylamide mix	2000	4000	6000	8000
1,0 M Tris (pH 8,8)	1300	2500	3800	5000
10 % SDS	50	100	150	200
10 % ammonium persulphate	50	100	150	200
TEMED	2	4	6	8

Table 21: Tris-glycine SDS-polyacrylamide gel electrophoresis: solutions for 5% stacking gel. Shown in grey are the normally used amounts.

Solution components	Component volumes [μ l] per gel volume of			
	1 ml	2 ml	4 ml	8 ml
5 %			1 gel	2 gels
H ₂ O	680	1400	2700	5500
30 % acrylamide mix	170	330	670	1300
1,0 M Tris (pH 6,8)	130	250	500	1000
10 % SDS	10	20	40	80
10 % ammonium persulphate	10	20	40	80
TEMED	1	2	4	8

Polypeptides separated by SDS-polyacrylamide gels were simultaneously fixed with methanol:glacial acetic acid and stained with Coomassie Brilliant Blue R250. The gel is immersed for several hours in a concentrated methanol/acetic acid solution of the dye and excess dye is then allowed to diffuse from the gel during a prolonged period of de-staining. The required solutions are given below:

Gel staining solution:

0.25	g	Coomassie Brilliant Blue R250
90.00	ml	methanol: H ₂ O (1:2 v/v)
10.00	ml	glacial acetic acid
<hr/>		
100.00	ml	

Gel stripping solution:	90 ml methanol:H ₂ O (1:2 v/v)	
	10 ml glacial acetic acid	
	<hr/>	
	100 ml	
450 ml methanol:H ₂ O (1:2 v/v)	200 ml methanol/H ₂ O/acetic acid	
50 ml glacial acetic acid	0.5g Coomassie Brilliant Blue R250	
<hr/>	<hr/>	
500 ml	200 ml staining solution	300 ml de-staining solution

450 ml methanol:H₂O (1:2 v/v) were mixed with 50 ml glacial acetic acid resulting in 500 ml of stripping solution. In 200 ml of this solution 0.5 g of Coomassie Brilliant Blue R250 were dissolved and well mixed. The solution was then filtered through a Whatman No. 1 filter to remove any particulate matter. The gel was immersed in at least 5 volumes of staining solution and placed on a slowly rotating platform for a minimum of 2 hours at room temperature. The stain was removed and saved for further use. After staining the gel was stripped by soaking it in the methanol/acetic acid solution without the dye on a slowly rocking platform for 4 - 8 hours, changing the solution up to three times.

The more thoroughly the gel was stripped, the smaller the amount of protein that could be detected by using this method. Destaining for 24 hours usually allows as little as 0.1 µg of protein to be detected in a single band. A permanent record of the gel was made, either by photograph/scanning or drying the gel under vacuum at 80°C. To dry the gels they were sealed between a foil and a filter: some water was put on a Whatman 3MM paper on the gel dryer. The gel was placed on the soaked Whatman paper and a foil was put on top of the sandwich (the foil was shrinking when getting wet, so a glass pipette was used to plane the foil). The gel dryer was sealed, and vacuum was applied. The lid was closed and the drying process was started by applying 80°C for 2 h. After drying the vacuum was removed and the gel, which was then attached to the 3MM paper, could be stored.

2.9.3 Semi-dry Protein Transfer for Western Blotting

For transferring the separated proteins from a SDS-PAGE to a nitrocellulose membrane the following materials and solutions were required:

Blotting Membrane (Immobilon™, PVDF-Membrane, pore size 0.45 µm) Millipore

SDS-PAGE (with pre-stained protein-marker)

Whatman paper (4 pieces with the size of the SDS-separation gel)

blotting chamber e.g. from BioRad (Trans-blot™ SD - Semi DRY Transfer cell)

Solution

blotting-, transfer-buffer:	25 mM Tris-buffer pH 8.3	(9.09 g / 3 l)
	192 mM glycine	(43.2 g / 3 l)
	15 % (v/v) methanol	(450 ml)

All preparation steps were done wearing gloves to prevent contaminations from the skin. Four pieces Whatman paper 3MM were cut to the size of the gel and pre-soaked in transfer buffer. The membrane (nitrocellulose filter) was also cut exactly to the size of the gel and incubated for 3 sec in to 100 % Methanol and kept permanently humid. First it was washed for 1 - 3 minutes into double distilled water and then equilibrated for another 10 minutes in transfer buffer. The SDS-PAGE was also incubated in transfer buffer for the same time. After equilibration the blot was assembled like shown in Figure 30:

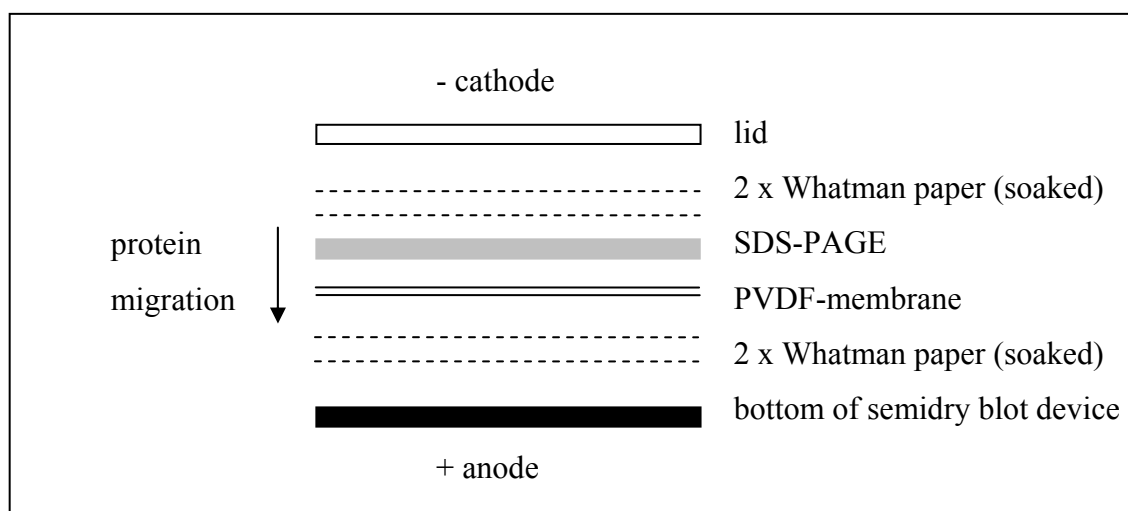


Figure 30: Blot assembly using a semi-dry transfer unit. The proteins migrate from the cathode to the anode. Consequently they are transferred from the SDS-PAGE to the PVDF membrane.

The resolving gel was put on the anode (+) of the transfer sandwich. On each side of the sandwich two soaked Whatman papers were used. All air bubbles had to be removed using a glass pipette, since they would affect the efficiency of the electro blotting. Proteins were electro-transferred from the gel on the membrane at room temperature. The current was kept at around 0.8 mA per 1 cm² of gel and the voltage at 15V. When the transfer was completed, the power was turned off and the layers of the sandwich were removed until the membrane was reached. The membrane was washed twice in double distilled water.

2.9.4 Immunological Detection of Proteins on Nitrocellulose Filters

To analyse the western-blot, the proteins had to be labeled with immunoglobulins. Just as proteins transferred from the SDS-polyacrylamide gel could bind to the nitrocellulose filter, so could proteins in the immunological reagents be used for probing. The sensitivity of western blotting depended on reducing this background of non specific binding by blocking potential binding sites with irrelevant proteins. The best and least expensive of several blocking solutions is non fat dried milk. After blotting and washing of the nitrocellulose membrane it was placed in a plastic container and 50 ml blocking solution (at least 0.1 ml of blocking solution per square cm of filter) was added. The required solutions are given here:

Tris buffered saline: (TBS-buffer)	Tris/HCl pH 7.5 NaCl	100 mM 150 mM
	store and use at 4°C	
Tween-TBS-buffer: (TTBS-buffer)	Tween20 0.1 % (v/v) in TBS store and use at 4°C	
Blocking solution:	TTBS + 5 % non fat dried milk (EC No. 2710453 Fluka for microbiology: 2.5 g/50 ml)	
Reaction buffer:	Tris/HCl pH 9.0 NaCl MgCl ₂	100 mM 100 mM 5 mM

Detection solution:	33 μ l	5-bromo-4-chloro-3-indolyl-phosphate (BCIP) (50 mg/ml in N,N-dimethylformamide)
	40 μ l	4-nitroblue-tetrazoliumchloride (NBT) (100 mg in 1.0 ml 70 % N,N-dimethylformamide)
	20 ml	reaction buffer

After blocking for at least 1 h with gentle agitation on a platform shaker the PVDF membrane was directly incubated with an antibody against the target protein. An appropriate dilution of the primary antibody was determined in pilot experiments. A clear band could be detected at a dilution of 1:1000 with almost no to very weak background. The primary antibody was diluted 1:1000 with blocking buffer and added to the nitrocellulose membrane in the tray. The membrane was incubated for 2 hours at room temperature with gentle agitation on a platform shaker. The incubation step with the first antibody could be increased to 18 hours and resulted sometimes in increased sensitivity of detection of the antigen. However the background of non specific binding also increased.

After incubation with the primary antibody the blocking solution and the antibody were discarded. The membrane was washed three times, ten minutes each, with at least 50 ml of TTBS buffer. The washing solution was discarded and the membrane was then incubated for one hour with the enzyme-coupled secondary antibody (1:1000; 3 μ l in 30 ml block-buffer), which had been diluted following the manufacturer's instructions. Following the incubation with the secondary antibody the membrane was washed again three times with 50 ml TTBS. Finally it was rinsed with TBS to remove residual Tween.

For detection of the enzyme coupled-secondary antibody 20 ml detection solution was used. After several minutes of incubation the precipitation at the protein antibody band became visible (violet colour). To stop the reaction the membrane was washed several times with water. After washing the membrane was dried, scanned and could be stored at room temperature.

2.10 Methods using DNA Microarrays

The following section will give some basic guidelines for working with DNA microarrays for expression profiling. Sample preparation, hybridization and detection will be discussed and a short introduction into data evaluation is given.

2.10.1 Labeling of aRNA/cDNA

Several different possibilities exist concerning labeling in microarray experiments, as recently reviewed by Brownstein, 2006. The most common ones are labeling methods using modified nucleotides which are incorporated in the aRNA/cDNA during reverse transcription or RNA amplification. The modification is in most cases a fluorescent dye which allows detection of labeled nucleic acid species by a fluorescence scanner. Depending on the labeling methods 0.5 to 20 µg of total RNA sample are required which equals around 100,000 to 5,000,000 hepatocyte cells. The labeling method used in this work was mainly the indirect labeling of amplified aRNA. For comparison and when sufficient material was available also direct and indirect labeling during the reverse transcription reaction was used.

2.10.2 Indirect Labeling of Amplified (a)RNA

Due to the limitation of sample material an amplification step (2.8.5) was often used to boost template levels from picograms to several micrograms without altering the complexity or composition of the original mRNA pool in the process. The labeling occurred during the so called *in vitro* transcription (IVT), a linear amplification with phage RNA polymerase (T7 RNA polymerase). In the version recommended by Eppendorf (MessageAmp™ II aRNA Kit from Ambion) and also the Affymetrix protocol (Dennise *et al.*, 2006) uses biotinylated nucleotides which are incorporated into the aRNA (2.8.5).

2.10.3 Direct Labeling During Reverse Transcription

The direct fluorescence labeling was done based on the DualChip[®] microarrays technical manual version 3.0. The design of the DualChip[®] microarray allowed one colour gene expression analysis, because the induced and control samples which were labeled identically could be hybridized on the two separated array fields on one slide. In this one colour experiments the two RNA samples to be analysed were labeled in two separate reverse transcription reactions using the same type of Cy3- or Cy5-labeled nucleotides in all reactions. Consequently all resulting cDNAs were labeled identically. For two-colour gene expression analysis, two RNA sets to be analysed were labeled in two separate reverse transcription reactions using two different kinds of labeled nucleotides (Cy3-dCTP and Cy5-dCTP). cDNA 1 derived from RNA 1 (control) was labeled with one dye and cDNA 2 derived from RNA 2 (treated sample) with the other dye. The two differently labeled cDNAs were mixed and the mixture was hybridized to one of the two arrays. Due to unequal performance of the two labels (Cy3 and Cy5), like different incorporation efficiency and quantum yields, it is strongly recommended to perform a dye-swap experiment using array 2. To do so, the labeling procedure for the two RNA samples to be compared, are performed vice versa with an exchange of the two labels.

The amount of starting material recommended to be used for one- or two-colour labeling with the direct fluorescence labeling protocol from Eppendorf was 20 - 25 µg of total RNA. First of all, components needed had to be thawed on ice: 100 mM dATP, 100 mM dCTP, 100 mM dGTP and 100 mM dTTP solution, Cy3- and / or Cy5-dCTP, Oligo(dT)₁₂₋₁₈ primer, RNase-free water, 5x First-Strand buffer, 0.1 M DTT, RNaseOut[™], SuperScript[™] II / III Reverse Transcriptase, RNase H, internal standard mix fluorescence, RNA sample.

The 100 mM dCTP solution was diluted with RNase-free water to a final concentration of 16 mM (1.6 µl 100 mM dCTP + 8.4 µl RNase-free water = 16 mM dCTP). Using this dilution a 10x Fluorophore-dNTP mix (Table 22) was prepared:

Table 22: 10x Fluorophore-dNTP mix for direct fluorescence labeling of samples for microarray analysis

Component	Volume	Final concentration in the 10x Fluorophore-dNTP mix
100 mM dTTP	1.5 μ l	5 mM
100 mM dGTP	1.5 μ l	5 mM
100 mM dATP	1.5 μ l	5 mM
16 mM dCTP	1.5 μ l	800 μ M
1 mM Cy3- or Cy5-dCTP	24.0 μ l	800 μ M
	30.0 μ l	

Subsequently the reaction mix was prepared on ice:

Reaction Mix 1:

Internal standard fluorescence-mix	2 μ l
Oligo (dT) ₁₂₋₁₈ primer (0.5 μ g/ μ l)	2 μ l
Total RNA sample	x μ l
ad. to final volume with RNase-free water	10 μ l

Reaction mix 1 was incubated at 70°C for 10 min., briefly centrifuged and chilled on ice for 5 min. Meanwhile the reaction mix 2 was prepared on ice, gently mixed and also centrifuged to collect everything at the bottom of the tube:

Reaction Mix 2:

5x First-Strand buffer	4.0 μ l
0.1 M DTT	2.0 μ l
10x fluorophore-dNTP mix	2.0 μ l
RNaseOut TM (40 U/ μ l)	1.0 μ l
	9.0 μ l

Reaction mix 2 (9 μ l) was added to reaction mix 1 (10 μ l) on ice. The solutions were well mixed, incubated for 5 min at room temperature and briefly centrifuged. 1.5 μ l SuperScriptTM II / III Reverse Transcriptase (200 U/ μ l) was added and mixed. After a short centrifugation step the reaction was incubated at 42°C for 90 minutes. After 90 minutes and a brief centrifugation step another 1.5 μ l SuperScriptTM II / III Reverse

Transcriptase (200 U/ μ l) was added and mixed. Following a 90 minutes incubation step at 42°C the reaction was heated to 70°C for 15 minutes. 1 μ l of RNase H (2 U/ μ l) was added and the mixture was briefly centrifuged and incubated at 37°C for 20 minutes. The reaction was terminated by heating to 95°C for 3 minutes. The samples were centrifuged and then placed on ice before proceeding to the purification.

To avoid bleaching the labeled cDNA was always protected from light. If two colour gene expression analysis was performed, both reactions (Cy3 and Cy5) were combined in one tube prior to purification. Unincorporated dNTPs, fluorescence dyes and primer were removed using the Qiagen PCR purification kit according to the manufacture's instructions. After purification the cDNA-containing solution was almost completely evaporated in a vacuum centrifuge at room temperature and the pellet was resuspended in 20 μ l RNase-free water. The purified, labeled cDNA could then be hybridized on the array or stored at -20 °C.

2.10.4 Indirect Labeling During Reverse Transcription

The indirect fluorescence labeling was done based on the DualChip[®] microarrays technical manual version 3.0. The basic principle of the indirect labeling is that instead of incorporating a modified nucleotide containing a fluorescent dye (as previously described); a biotin-modified nucleotide is used. The visualisation is performed later on by detecting this modified nucleotide by a biotin specific antibody-Cy3-conjugate. So the fluorescence label is applied to the samples after reverse transcription and hybridisation. As the samples are both labeled with biotin-nucleotides they have to be hybridized on different arrays like for the direct one colour labeling protocol (RNA 1 = control = array 1 and RNA 2 = treated = array 2). As there is no bias due to different dyes used for labeling, a dye-swap experiment is not necessary.

Due to increased sensitivity the recommended amount of starting material to be used for one colour indirect fluorescent labeling protocol from Eppendorf was 5 - 10 μ g of total RNA. For all samples with limited amount of RNA an amplification step including labeling was performed (2.8.5) based on the MessageAmp[™] II-Biotin Enhanced Kit" from Ambion.

For the indirect labeling without amplification the needed components were thawed on ice: 100 mM dATP, 100 mM dCTP, 100 mM dGATP and 100 mM dTTP solution, Biotin-11-dCTP solution (5 mM) and Biotin-11-dATP solution (5 mM), Oligo (dT)₁₂₋₁₈ primer, RNase-free water, 5x First-Strand buffer, 0.1 M DTT, RNaseOut™, SuperScript™ II / III Reverse Transcriptase, RNase H, internal standard mix fluorescence, RNA sample.

The 100 mM dCTP and dATP solutions were diluted with RNase-free water to a final concentration of 16 mM (1.6 µl 100 mM dNTP + 8.4 µl RNase-free water = 16 mM dNTP). Using this diluted dNTP solutions a 10x Biotin-dNTP mix (Table 23) was prepared:

Table 23: 10x Biotin-dNTP mix for indirect fluorescence labeling of samples for microarray analysis

Component	Volume	Final concentration in the 10x Fluorophore-dNTP mix
RNase-free water	19.2 µl	
100 mM dTTP	2.0 µl	5 mM
100 mM dGTP	2.0 µl	5 mM
16 mM dATP	2.0 µl	800 µM
16 mM dCTP	2.0 µl	800 µM
5 mM Biotin-11-dCTP	6.4 µl	800 µM
5 mM Biotin-11-dATP	6.4 µl	800 µM
	40.0 µl	

The reaction mix was prepared on ice:

Reaction Mix 1:

Internal standard fluorescence-mix	2 µl
Oligo(dT) ₁₂₋₁₈ primer (0.5 µg/µl)	2 µl
Total RNA sample	x µl
ad. to final volume with RNase-free water	10 µl

In the next step the reaction mix 1 was incubated at 70°C for 10 min, briefly centrifuged and chilled on ice for 5 min. Meanwhile the reaction mix 2 was prepared on ice, gently mixed and also centrifuged to collect everything at the bottom of the tube:

Reaction Mix 2:

5x First-Strand buffer	4.0 μ l
0.1 M DTT	2.0 μ l
10x fluorophore-dNTP mix	2.0 μ l
RNaseOut™ (40 U/ μ l)	1.0 μ l
	9.0 μ l

Reaction mix 2 (9 μ l) was added to reaction mix 1 (10 μ l) on ice. The solutions were well mixed, incubated for 5 min at room temperature and briefly centrifuged. 1.5 μ l SuperScript™ II / III Reverse Transcriptase (200 U/ μ l) was added and mixed. After a short centrifugation step the reaction was incubated at 42°C for 90 minutes. After 90 minutes and a brief centrifugation step another 1.5 μ l SuperScript™ II / III Reverse Transcriptase (200 U/ μ l) was added and mixed. Following a 90 minutes incubation step at 42°C the reaction mix was heated to 70°C for 15 minutes. 1 μ l of RNase H (2 U/ μ l) was added and the mixture was briefly centrifuged and incubated at 37°C for 20 minutes. The reaction was terminated by heating to 95°C for 3 minutes. The samples were centrifuged and then placed on ice before proceeding to the hybridization.

2.10.5 Hybridization and Washing

The hybridization mix was prepared in the following order:

Labeled cDNA (Biotin or pooled Cy3/Cy5)	20.0 μ l
Biotin-HybControl (for indirect labeling protocol)	
Fluorescence-Hyb control (direct labeling)	10.0 μ l
RNase-free water	20.0 μ l
HybriBuffer A	10.0 μ l
HybriBuffer B*	40.0 μ l
Final volume	100.0 μ l

* If a white precipitate appeared, the HybriBuffer was heated for 5 min to 60°C.

The solution was slightly mixed by pipetting and centrifuged briefly. The mixture was heated for 5 minutes to 60°C. To apply the hybridization mix on the array, the solution was reared in a 100 µl pipette tip. The filled tip was then vertically inserted into the injection port of the hybridization frame. Avoiding the injection of air bubbles, the 100 µl were slowly dispensed until the complete array was filled with solution and a small liquid excess appeared at the opposite port. If some liquid had spilled out of the ports they were dried, not touching the window of the hybridization frame. The ports were sealed using a sticky aluminium sealing pad by putting it on top of the chamber and applying slight pressure along the wall. Immediately after sealing the DualChips® were placed in a Thermomixer and incubated overnight (12 - 16 hours) at 60°C with a mixing frequency of 1400 rpm.

The post hybridization procedure for direct labeling consisted only of washing, whereas during indirect labeling an additional detection step (2.10.6) was necessary. Two washing buffers were prepared by diluting the UniBuffer (supplied in the kit). For wash buffer one 50 ml UniBuffer were mixed with 1950 ml distilled water and 2 ml of Tween 20 was added. Wash buffer two consisted of 50 ml UniBuffer in 1950 ml distilled water without Tween. The buffers were mixed prior to use. The washing dishes were filled with 200 ml of wash buffer one. After the overnight hybridization step, the slides were removed from the thermoblock and incubated for 2 minutes at room temperature. Both hybridization frames were removed with fine-tip tweezers. The slide was immediately immersed into wash buffer one in the washing dish. It was important to avoid drying on the slide to prevent increased background fluorescence. If the adhesive part of the hybridization frame stayed on the slide it was nevertheless immersed in wash buffer. The glue was removed later on during the washing process. The DualChip® was then washed twice in wash buffer one for 2 minutes at room temperature (agitation was not needed). If necessary the remaining adhesive part of the hybridization frame was removed with tweezers. The slide was incubated again for 2 minutes in wash buffer one followed by two washing steps for 2 minutes in wash buffer two. Excess of liquid was removed by gently knocking the edge of the slide on a paper towel before the chips were dried by centrifugation in a special adapter ("CombiSlide" adapter, Eppendorf) for 5 min at 600 rpm or using a nitrogen stream

moving the air flow downward from the sticker side. After drying the slides were ready to be scanned.

2.10.6 Detection of Indirectly Labeled Targets

For indirectly labeled samples an additional labeling step was necessary. In addition to the two washing buffers a blocking buffer (1 ml per chip) and a Cy3-conjugate-antibody solution had to be prepared. To prepare 10 ml blocking buffer a 4-fold dilution from the UniBuffer was done (2.5 ml UniBuffer with 7.5 ml distilled water). The content of one tube of blocking reagent powder (skim milk) was dissolved in the diluted UniBuffer. The solution was mixed by slightly shaking until it was completely dissolved. The antibody detection solution was prepared by diluting the conjugate 1:1000 in blocking buffer.

After removing the hybridization frames from the slide and the subsequent two washing steps in washing buffer one (2.10.5) the detection was performed on an antibody incubation foil. 800 μ l of diluted conjugate-Cy3 was pipetted on the foil. Excess of liquid was removed from the slide by gently knocking one edge of the slide on a paper towel. The slide was put, sticker upside down, arrays facing the drop of conjugate without any pressure on the slide and incubated for 45 min at room temperature. After incubation excess of conjugate was removed by knocking the slides slightly on a paper towel. The DualChip® was washed 5 times in wash buffer one for 2 minutes at room temperature. To remove traces of wash buffer, the slides were washed twice in distilled water for 2 minutes at room temperature.

The array was dried as described previously (2.10.5) using a special adapter. The arrays were then ready to be scanned.

2.10.7 Scanning of Arrays

A microarray scanner (Perkin Elmer) was used to detect the hybridized Cy3-labeled cDNA. In order to reach the full dynamic range of the system and to increase the number of genes the DualChip® was scanned with three different settings for the photomultiplier tube (PMT). An intermediate PMT setting was first chosen to analyse

genes which are moderately expressed. High expression genes which would give at this intermediate PMT a saturated signal can be accurately quantified at a low PMT setting. For low expressed genes a high PMT setting was chosen to yield in a quantifiable signal. In most of the cases the PMT settings for the scanner were 100 %, 70 % and 50 %. As a rule of thumb it can be assumed that the resulting image of the array scanned at the lowest PMT setting, should have no saturated signals. The excitation for Cy3 is at 550 nm and the emission wavelength is at 570 nm. For Cy5, the excitation wavelength is 649 nm and the emission is 670 nm. Scanning was performed with 50 μm resolution for an overall preview scan and 5 μm resolution was used for the final scan, which was quantified in the following step. The images were stored as 16 bit tif-file and false colour jpg image.

2.10.8 Extraction of Raw Data and Picture Analysis (Quantification)

The quantification of the scanned image of the array was done using a quantification protocol programmed with the software scan-array express from Perkin Elmer. The array specific data (grid size, number of rows, columns, etc.) was obtained by a gal file provided in the Eppendorf kit. Relevant parameters are given in Table 24.

The grid was adjusted with the spots and the spot finding algorithm was started. The quantification method used was adaptive threshold. The advantage of the adaptive threshold against the conventional one is, instead of using a global threshold for all pixels, the adaptive operator changes the marginal value dynamically over the image. This more sophisticated version of an adaptive threshold can accommodate changing lighting conditions in the image, e.g. those occurring as a result of a strong illumination gradient, shadows or other background effects.

Adaptive threshold used the scanned 16 bit tif-file as input and outputs a binary image representing the segmentation. For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold it is set to the background value, otherwise it is used for quantification. The assumption behind the method is that smaller image regions are more likely to have approximately uniform illumination, thus being more suitable as threshold. The image is divided into an array of overlapping sub images and then the optimum threshold for each sub image is searched by investigating

its histogram. The threshold for each single pixel is found by interpolating the results of the sub image. Position of each spot, allocated gene, spot foreground intensity value and background intensity were stored in a comma separated value file (CSV). The last line of this file had to be erased for further processing in the DualChip® software. The value table was stored as txt-file and further analysed using Microsoft Excel and the DualChip® evaluation software from Eppendorf.

Table 24: Parameters that describe the quantification protocol for the Eppendorf DualChips®

Parameter	value
No. of sub arrays, rows	1
No. of sub arrays columns	1
Pin spacing, H (mm)	4.5
Pin spacing, V (mm)	4.5
Rows of spots per sub array	27
Columns of spots per sub array	21
Spot spacing H (μm)	643
Spot spacing V (μm)	500
Spot diameter (μm)	200
Quantification method	Adaptive Threshold
Max Spot diameter (μm)	240
Inner background diameter (μm)	527
Outer background diameter (μm)	814

2.10.9 Evaluation of DNA Microarrays and Data Analysis

Plotting data and identification of differentially expressed genes without normalization of the data could be done in Excel to get a first impression of the quality of the obtained data and the experiment. This was achieved by subtracting the background intensity from the signal intensity and plotting the background corrected signals of sample one against the background corrected signals of sample two on a log scale (Figure 31).

Not differentially expressed genes were lying on the bisecting line while up-regulated genes were above and down-regulated genes below this line. Although the control spots could be validated whether positive probes had a high signal and negative probes a low signal. However a more detailed and sophisticated analysis was done using on the one hand the Eppendorf DualChip® Software from Eppendorf for the Eppendorf DualChips® and on the other hand the Analyst from Genedata for the Affymetrix Chips.

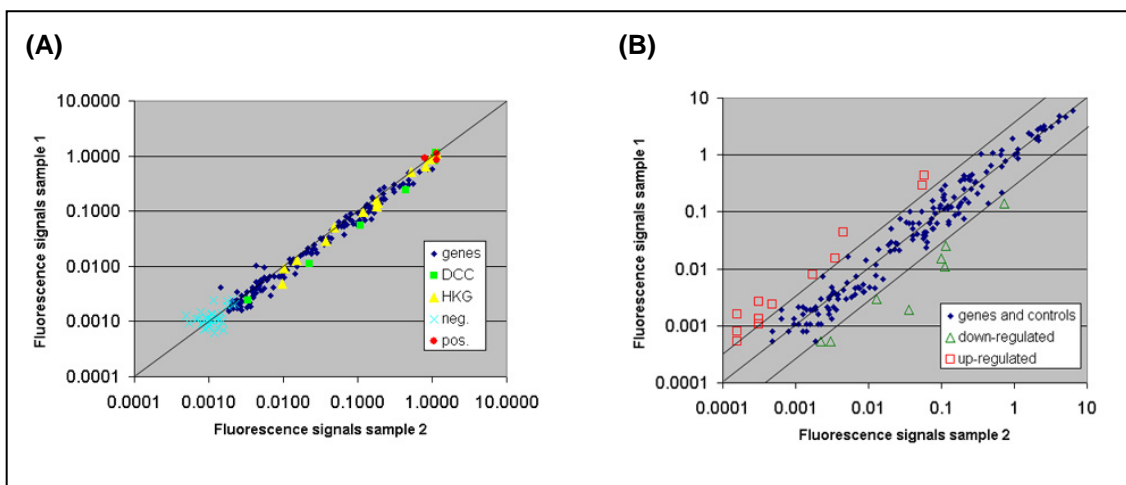


Figure 31: Scatter Plot of expression raw data from a DNA microarray experiment. (A) Colour coded data from a self-self hybridization experiment. The same samples were hybridized on two arrays and plotted against each other. As a result no differential expression is observed, as all spots are on the bisecting line. Genes of interest are shown in blue (◆). The detection curve concentration which is well distributed on the whole dynamic range is given in green (■). Housekeeping genes are shown as yellow triangles (▲). Finally controls are shown as light blue "x" for the negative controls and red dots (●) for the positive controls. Part (B) shows a scatter plot where a test and a reference sample were hybridized. The differentially expressed genes can be identified by lying apart from the bisector line. Up-regulated genes can be seen above the line as red squares (□) and down-regulated genes are shown by green triangles (△).

2.10.10 Analysing Sub-Genome Data using the Eppendorf DualChip® Software

The workflow implemented in the software included capabilities for the extraction of gene expression intensities from quantification data obtained from (2.10.8), the normalization of these raw data, the identification of significantly, differentially expressed genes of an experiment in comparison to a reference and the visualization and saving of the results. As the arrays were scanned at different PMT settings to

increase the dynamic range of the system it was also important that the software was capable to allow simultaneous analysis of arrays scanned at different scanner settings. Additionally the software is designed for the comparison of several replicates of gene expression profiles obtained by the data analysis workflow shown in Figure 32.

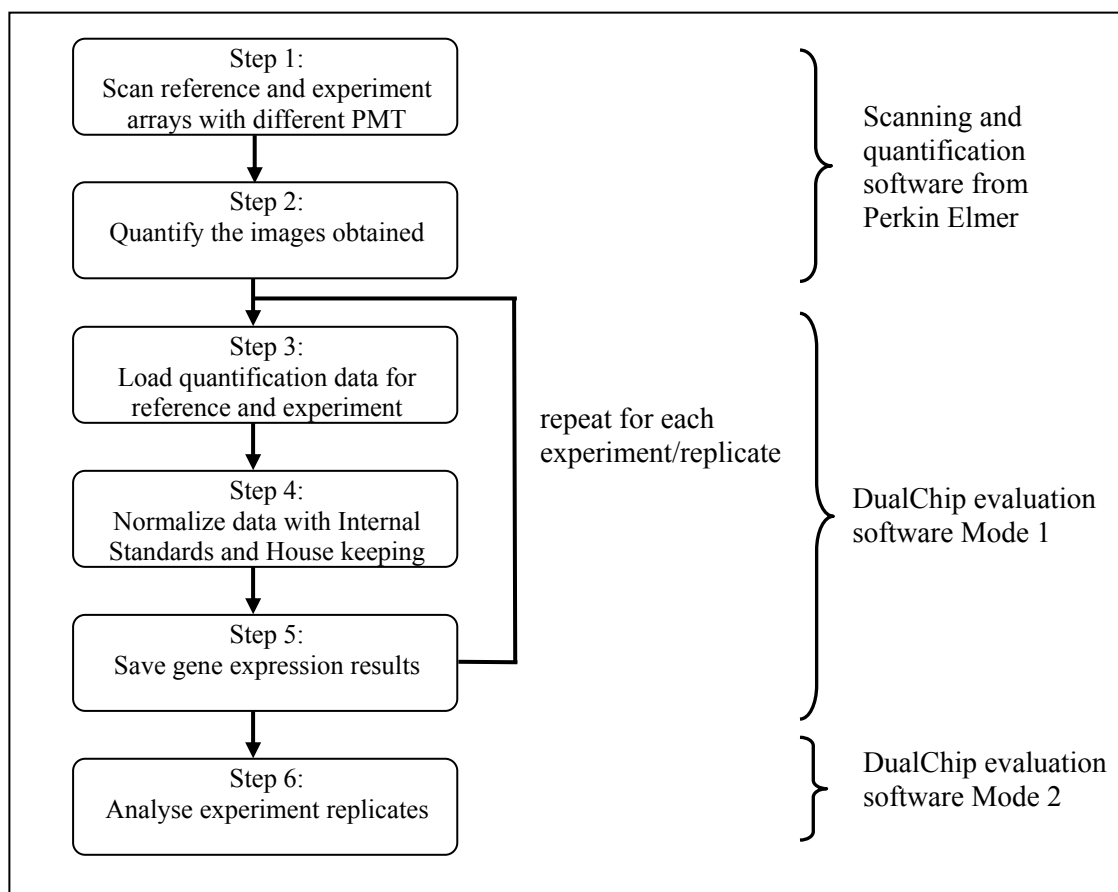


Figure 32: Flowchart showing a classical analysis of gene expression experiments using Eppendorf DualChips® and the DualChip® evaluation software (version 1.1). The experiment consists of 6 steps: first of all scanning, quantification, loading of data in the DualChip® evaluation software for analysis, normalization and identification of expression results is performed. In a second part the expression results are used for analysing replicates of the same experiment.

The extensive set of control probes printed on the DualChip® microarray includes positive hybridization controls, negative hybridization controls, positive detection concentration curve (DCC), negative detection controls, an internal Standard and a set of housekeeping genes. This control set allows verification of cDNA/aRNA synthesis efficiency and hybridization as well as signal linearity. Furthermore these controls are intended to control and correct the possible variables occurring during the experiment. A summary of all controls on the DualChips® and their purpose is given in Table 25.

Table 25: Kind of controls on the DualChips® from Eppendorf and their purpose

control name	type of molecule / step	usage
Internal Standard (mix of spiking RNAs)	in-vitro transcribed poly-adenylated RNA (Lycopersicon esculentum) spiked into RT	reverse transcription- or RNA amplification-, hybridization control and normalization
Positive hybridization controls	biotinylated DNA spike into the hybridization mix	hybridization control, grid alignment (indirect labeling)
Negative hybridization controls	Plant DNA on DualChips®	detection of non specific hybridization (subtracted from signal)
Negative detection controls	buffer on DualChip®	check of quality of detection
Positive detection concentration curve	biotinylated capture probe on DualChip®	detection control and for grid alignment (indirect labeling)
Housekeeping genes (HKG)	cDNA / aRNA	normalization

The controls which enable the validation of the different steps of the analysis are given in the following table (Table 26):

Table 26: Interpretation of control spots to validate array results: present ("+" = positive signal) or missing ("- = negative signal"). Based on the evaluation of the control spots it was possible to detect whether errors occurred during cDNA/aRNA synthesis, hybridization or detection of signals.

positive detection control*	Internal Standards	Positive hybridization controls**	Interpretation
+	+	+	All steps are successful
+	-	-	The cDNA / aRNA synthesis or hybridization step has failed
+	-	+	The cDNA / aRNA synthesis does not occur properly
-	-	-	The detection step has failed

* only valid for indirect labeling protocol

** for direct labeling only when using the optional hybridization control

The quantified data were loaded into the DualChip® Software using the following settings for the parsing parameters: delimiter symbol ","; first row of data "63" or "61" depending of the DualChip® Version; gene ID column "6"; signal column "15" and background column "19" for median signal and mean background values or signal column "16" and background column "18" for mean signal and median background. If the background was very homogenous a mean background calculation could be used, however in most cases it was more appropriate to use the median background.

After import of the raw data, replicates had to be merged. Potential errors such as bubbles or dust may alter the signal intensity of spots. The presence of spot replicates provides valuable quality information and allows the performance of statistical analysis and the correction of the gene data within the experiment. Having triplicates, spots showing an abnormal variation of signal intensity are considered as outliers. To be able to find and remove such faulty spots the coefficient of variation (CV = standard deviation/mean) of the replicate signal intensities had to be determined using the following rules. If the CV was higher than 0.7 and one replicate signal intensity is above 1000, the replicate with the highest deviation from the mean was removed. Due to the greater variability of low signals, the acceptance range for those genes is chosen to be broader. If the CV was higher than 0.3 and one signal was above 10000 again the replicate with the highest deviation from the mean was removed. Using these rules it was possible to discard potential outliers. Following this "spot flagging", the mean of remaining replicates (at least two replicates) was calculated.

Before starting the evaluation of differential gene expression the data were validated. Spots which showed fluorescence intensity that was not in the linear range of measurement, e.g. either saturated or very low signal intensities were sorted out. The validation process used the following criteria. If the intensity of the spot (S) was above 50000 fluorescence units the signal intensity of the spot was regarded as saturated. For an intensity S which was equal or less than its local background multiplied by two, the signal intensity of the spot was designated as too low ("not detected"). This could be expressed by the formula 5:

$$S \leq B \cdot 2 \quad (5)$$

An additional rule was used to compare the intensity of each spot to the intensity of the negative hybridization control spot after having subtracted their corresponding local background. If the intensity of the spot ($S - B$) was less than the mean intensity of negative hybridization controls (NHCS - B) plus the standard deviation of the signal intensity of the negative hybridization controls (minus background) multiplied by 2, the signal intensity of the spot is designated as too low ("not detected").

This can be expressed in the following formula 6:

$$(S - B) \leq [(NHCS - B) Mean + ((NHCS - B) Std dev) \cdot 2] \quad (6)$$

In other cases the spot was designated as quantifiable. The validation of mean gene intensity (mean of replicates) was performed in the same manner as the validation of single replicate intensity. A gene is designated as quantifiable when all the selected replicates are quantifiable. If not, the gene was designated as too low or saturated according to the replicates validation.

In the next step, gene expression ratios consisting of the signal intensities of experiment and reference are calculated for each gene. If the signal intensity for both the experiment and the reference was quantifiable, the ratio was termed "quantitative". If only one intensity was quantifiable (experiment or reference), the ratio is defined as "qualitative". If neither intensity is quantifiable the ratio is designated as being "not in linear range", as shown in Figure 33.

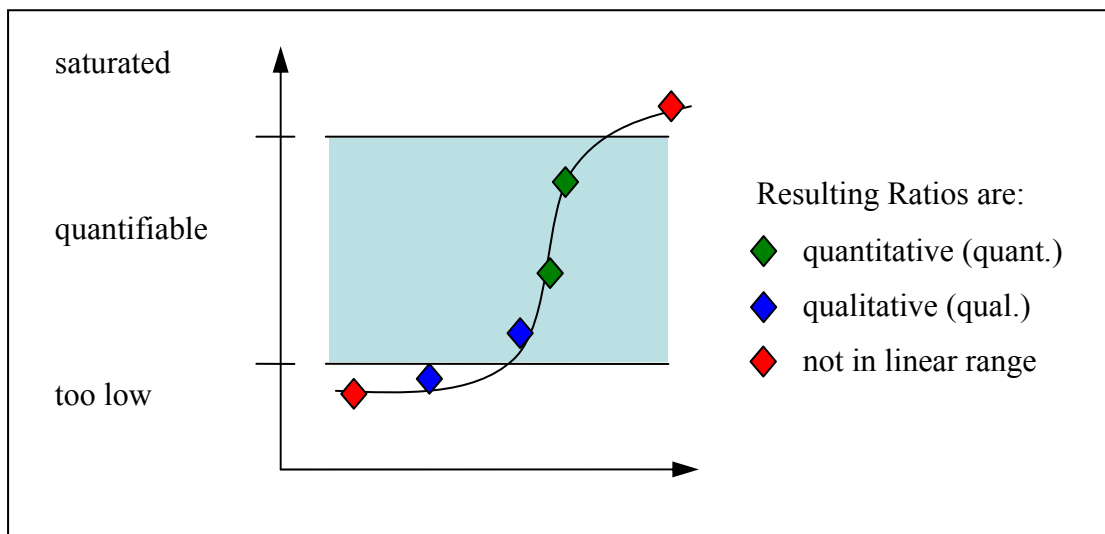


Figure 33: Definition of quantitative, qualitative and data which is not in linear range. If all replicates are in the quantifiable range that data is designated as quantitative (green). For measurements where one measurement is outside and one in the quantifiable range of the experiment it is qualitative (blue). If all measurements are outside the quantifiable range the data is marked as "not in linear range" (red).

Normalization was done using a two step normalization procedure based on an internal standard which was spiked into the sample and on 16 housekeeping genes. The DualChip® was divided into six zones, each containing two different concentrations of

internal standard (IS) capture probes. This IS capture probes correspond to differentially concentrated RNA spikes (low and high) of the standard. Using the quantifiable internal standard intensity a local normalization factor could be calculated. It is based on the intensity ratios of the internal standard in the reference and the induced samples. With this local normalization factor the ratios for all genes between experiment and induced samples were corrected.

Due to the fact that the internal standard is an exogenous RNA spike it does not reflect the purity, quality and amount of sample RNA to be analysed. Therefore a second normalization step based on expression level of housekeeping genes was performed. A housekeeping normalization factor is calculated using the mean of quantifiable housekeeping gene ratios.

Finally the statistical test for significance of gene expression ratios had to be performed. This test determines whether different gene expression is significant or due to random between reference and induced sample. The model used assumes the intensities to be Gaussian, independently distributed with a constant coefficient of variation. Based on these assumptions, the coefficient of variation and the confidence intervals were calculated.

Housekeeping genes which were assumed to be similarly expressed in the control samples and the induced samples were used to determine the coefficient of variation. The variation of the housekeeping genes should therefore be similar to the variation of non regulated genes, so the ratio between sample and control should be around one. The significance of an observation was determined by the confidence interval. Ratios outside the 95 % confidence interval were statistically significant and ratios outside the 99 % confidence interval were statistically highly significant.

Furthermore the ratios between sample and control were grouped into quantitative and qualitative results. Ratios based on two quantifiable signal intensities were considered as quantitative. Ratios involving one low or saturated signal were considered as qualitative. If none of the signals was in the quantifiable range (both signals saturated or too low) the ratio was designated as "not in linear range" (see Figure 33).

As mentioned before quantification was only possible if the signal intensities were in a quantifiable range. As the expression range of different target genes was quite broad, from very low expressed genes to very high expressed genes, it was advisable to scan

the arrays with different settings for the photomultiplier of the scanner (PMT) to increase the dynamic range of detectable genes. Low expressed genes could be detected with a high PMT setting and high expressed genes which would have a saturated signal could be detected using a low PMT setting. During the combination of ratios generated at different PMT settings, quantitative results overruled qualitative results and qualitative results overruled "not in linear range" results.

Finally the "overall ratios" generated, which were condensed from the ratios at different PMT settings could be plotted as bar charts and stored as jpg or text files for further comparison.

2.10.11 Analysing Full-Genome Data Using "Refiner" and "Analyst" from Genedata

Two software packages from Genedata for data processing of the full genome microarrays from Affymetrix were used: Refiner and Analyst. Refiner was used for loading of unprocessed raw data, detection and masking of defective regions, detection and correction of gradients and distortions, condensing (GC-RMA) of probes, generation of quality- and processing reports and storage of the data in the database.

The Analyst software package allows sophisticated statistical analysis and flexible visualization of the data containing fold-change analysis and marker identification by different clustering methods.

The raw data from an Affymetrix microarray experiment contains all information in a so called archive file (DDT). Using a data conversion tool from Affymetrix this DDT files were transformed into "cel" files containing the probe intensities, which were used as input for the Refiner workflow to import the data.

An overview of the workflow used in Refiner for loading the data and performing quality control and condensing is given in Figure 34.

The condensed data was normalized and stored in the database. In the next step the "Analyst" from the Genedata software package was used to identify differentially expressed genes. First candidates were determined by a fold change analysis.

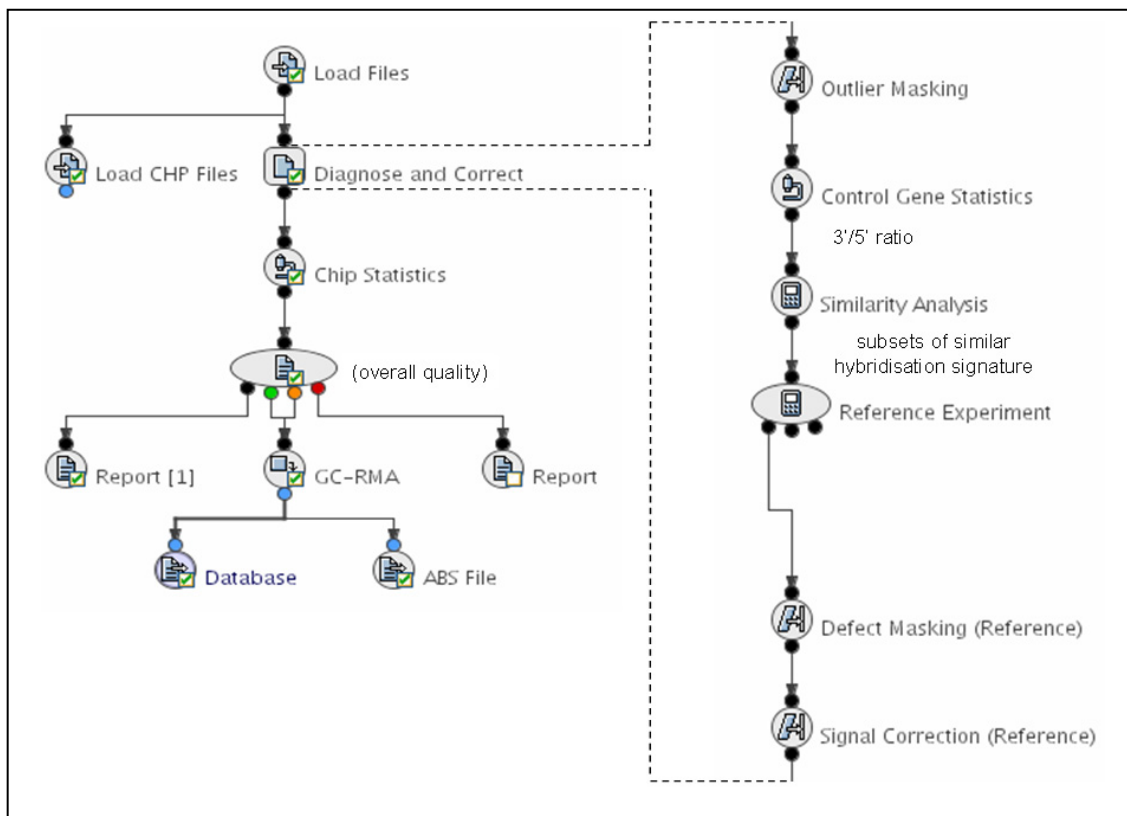


Figure 34: Workflow (GD AFFX Correction with Reference - Affymetrix2007.xml) for loading raw data from Affymetrix full genome microarrays using Refiner from Genedata. The data is loaded from cel files (containing the signal intensities for the probes) and analysed by several diagnosis procedures (like "outlier masking", determining 3'/5' ratios and "similarity analysis"). The arrays are grouped by the "overall quality" activity into all (left), good and medium (middle) and bad (right) ones. Reports for all and the bad chips are generated. For the arrays with good and medium quality the signals for the different probes of the same target genes are condensed (GC-RMA) and stored to the database.

Genes showing a fold change larger than 2 were tested for their statistical significance in the different biological samples using a standard (two sample) t-test based on the following formulas:

Let $x_{g,i}$ $1 \leq g \leq 2, 1 \leq i \leq n_g$ be the expression values in logarithmic form. The t-test is computed from the means of the expression values in one group (e.g. test or reference)

$$\bar{x}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} x_{g,i} \quad (7)$$

and the sample variances in the group

$$s_g^2 = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (x_{g,i} - \bar{x}_g)^2. \quad (8)$$

Its test statistics compares the differences in group means to an average standard deviation s_{12} ,

$$t = \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \frac{x_1 - x_2}{s_{12}}, \quad (9)$$

where the square of the average standard deviation is obtained as the weighted sum of the group variances

$$s_{12}^2 = w_1 s_1^2 + w_2 s_2^2 \quad (10)$$

with weights

$$w_g = \frac{n_g - 1}{n_1 + n_2 - 2}, \quad \sum_{g=1}^2 w_g = 1. \quad (11)$$

In terms of expression values, the test statistics takes the form

$$t = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sum_{g=1}^2 \sum_{i=1}^{n_g} (x_{g,i} - \bar{x}_g)^2}}, \quad (12)$$

where the dominator is the root of the total sum square of differences to the respective group means.

Based on one example comparing all test (Rifampicin) and control (DMSO) arrays from all biological (patients) and technical replicates, the whole procedure of identifying and testing will be displayed.

An often used graphical representation of gene expression levels in two groups to be compared is the scatter plot (Figure 35). The signal intensity of two groups is plotted along the x- and y axis and the ratio values are plotted as a distance from the diagonal.

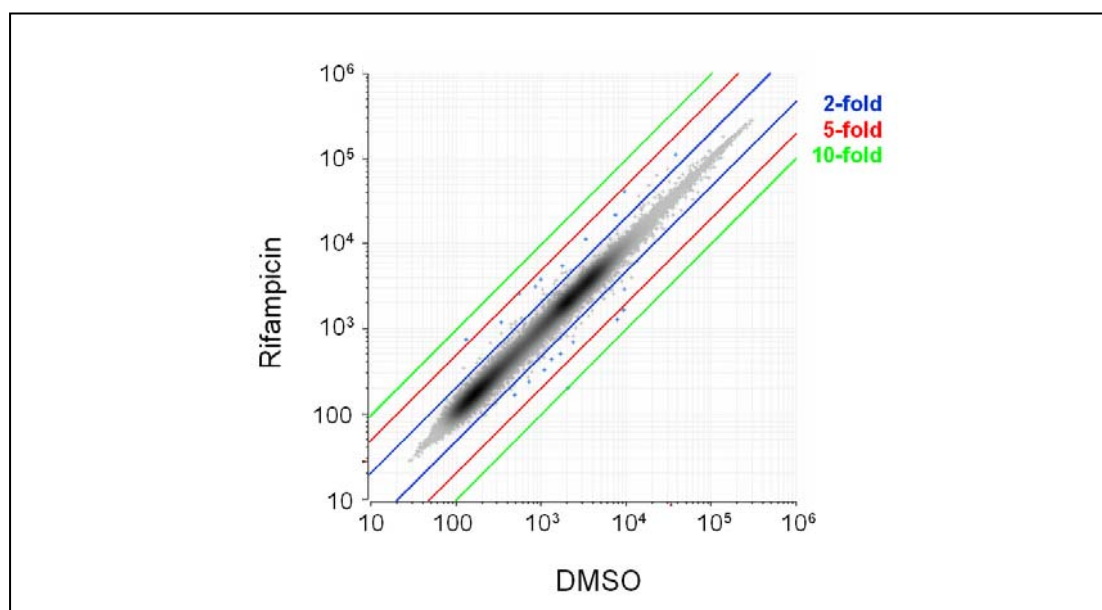


Figure 35: Scatter plot of all probes on all Affymetrix arrays used in this study. Plotted on the y-axis are the intensities representing the mRNA level of all probes (spots) of the Rifampicin chips. On the x-axis are the corresponding intensities for the control (DMSO) chips assigned. Genes showing no differential expression are on the bisecting line. If the genes are differentially expressed 2-fold they can be found outside the blue lines. Genes with a 5-fold difference between Rifampicin and DMSO are outside the red lines and with a 10-fold difference outside the green line. Genes which are upregulated in the Rifampicin treated cells compared to the DMSO cells are shown above the bisecting line and genes which are downregulated can be found below.

In the subsequent step all genes showing a two fold regulation were selected as shown in Figure 36. These genes were combined in one group (2-fold differentially expressed). The second information which can be obtained from a scatter plot, whether a gene is up- or down regulated was applied to the identified differentially expressed genes. Therefore genes lying above the bisecting- or identity line were selected and combined to one group (up-regulated). A second group was generated including all genes below the identity line (down-regulated). Before getting a list of 2-fold up- and down-regulated genes as final results, a t-test for statistical significance was performed (formula 12).

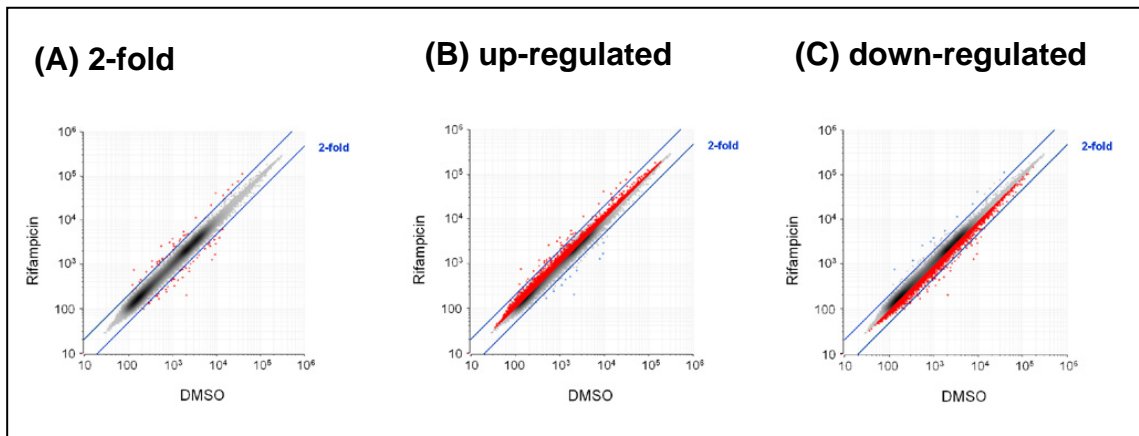


Figure 36: Scatter plots showing 2-fold regulated as well as up-regulated and down-regulated genes of Rifampicin and DMSO samples on Affymetrix arrays. In (A) the red dots represent genes having a fold change of 2 or larger. In part (B) all genes having a higher expression in Rifampicin samples compared to DMSO samples are shown in red and in (C) the similar labeling as in (A) is done for the down-regulated genes.

93 items (probes corresponding to genes) were identified to be two fold differentially expressed over all samples between Rifampicin and DMSO samples. Groups were generated which include all genes meeting the statistical criteria ($p < 0.01$ and $q < 0.05$). Finally the groups were connect using binary logical operators and are displayed here as Venn diagrams (Figure 37).

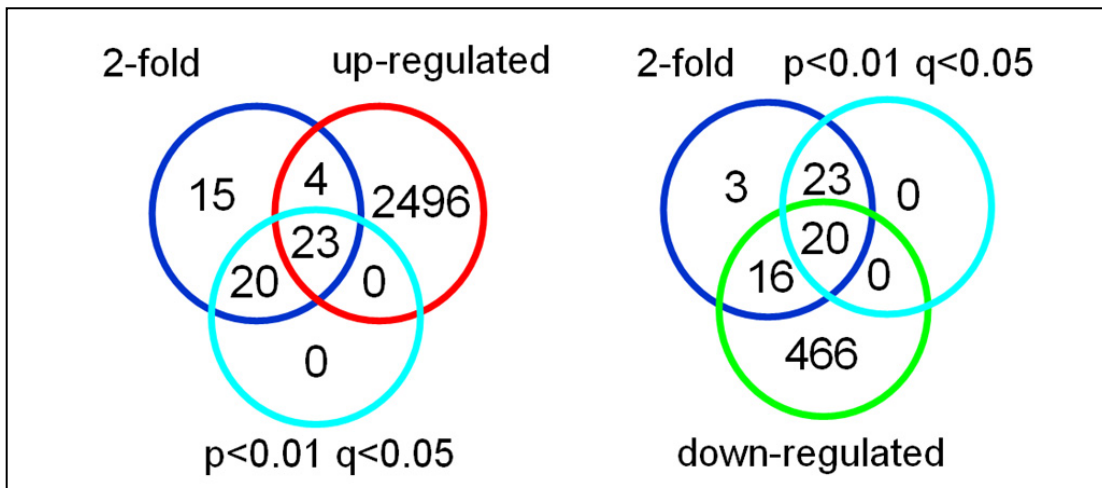


Figure 37: Venn diagram displaying relationship between 2-fold differentially expression, statistical significance of the observation and up- (left) or down-regulation (right).

Introduced 1881 by the British philosopher and mathematician John Venn (1834-1923) the Venn diagrams are illustrations showing all of the possible mathematical or logical relationships between groups.

As illustrated in the Venn diagrams (Figure 37) as dark blue circles 62 genes (derived from the 93 probes) showed an at least 2-fold difference in Rifampicin samples compared with control samples (DMSO). Shown in light blue are the 43 genes matching the statistical threshold of $p < 0.01$ and $q < 0.05$. Up-regulated genes (2,523) are marked with a red circle and down-regulated genes (582) with a green circle. By this visualization the intersection of all circles allows to determine the genes which are two fold differentially expressed, meet the statistical relevance and are up- or down-regulated. From these 43 genes, sequences coding for hypothetical proteins and unknown function were removed. The finally identified 17 significant 2-fold up-regulated and the 14 significant 2-fold down-regulated genes are given in Table 40 and Table 41 in the results part of this work.

To structure the data several clustering methods were applied. A good overview of clustering microarray methods is given by Gollub and Sherlock, 2006. Clustering is a multivariate classification method categorizing data based on the similarity of data points (gene expression levels, ratios). The clustering algorithms as well as the output rely on distance measurements between different data points from a set of data. As there are several distance functions available, some basic criteria for their selection will be given.

The expression level of two genes (i) and (j) was measured. How can the distance (d) between them be determined? The distance between two genes can be positive, meaning that if the genes i and j are different, the distance between them must be positive. If the genes show the same level of expression then the distance must be zero (as shown in formula 13):

1. $if(i \neq j), d(i, j) > 0$
2. $if(i = j), d(i, j) = 0$

(13)

The distance can also be symmetric (formula 14). In this case the distance from i to j is the same as the distance from j to i . That is for any two points i and j ,

$$d(i, j) = d(j, i) \quad (14)$$

The distance between two genes can also satisfy the triangle inequality (formula 15) meaning that the distance between two points can never be more than the sum of their distances from a third point. That is, for example for the three points i, j, z ,

$$d(i, j) \leq d(i, z) + d(z, j) \quad (15)$$

Common distance functions for gene expression pattern estimation are the Euclidean distance, Mahalanobis squared distance, Minkowski distance and Manhattan distance. Euclidean distance between two vectors representing expression values is just the distance in space between the two end points defined by those vectors. It is sensitive to the direction and magnitude of the vectors. These distance function is appropriate for variables that are uncorrelated and have equal variances, allowing reflection of dissimilarity between the values of two gene expression levels. The Euclidean distance is calculated as (formula 16)

$$\sqrt{\sum_{i=1}^n (x_{i,k} - x_{j,k})^2} \quad (16)$$

Mahalanobis squared distance (MSD) gives a one dimensional measure of how far a gene expression value is from a certain functional class. This distance measure is useful to identify genes having an abnormal behaviour. The Minkowski distance is especially useful when it is known the some functional groups are more important than others in a particular microarray experiment as it associated weights to each gene. This might help in identifying groups of genes which dominate others. Using the Manhattan distance, also known as "city block" distance, gene expression values are partitioned into homogenous rectangular blocks, based on formula 17, giving a good overview over genes sharing similar expression levels. The Manhattan distance between two points can be thought of the path one would have to follow between two addresses in an urban downtown, making right angle turns. The distance is calculated from the sum of absolute values of these orthogonal lines of the journey (as indicated in formula 17) rather than as the familiar sum of squares of Euclidean distance (formula 16). As each element of the vector is weighted linearly rather than quadratically the Manhattan distance is less sensitive to outlier values.

$$d_{i,j} = \sum_{k=1}^q |x_{i,k} - x_{j,k}| \quad (17)$$

In equation 17 q is the number of variables and $x_{i,k}, x_{j,k}, k=1, \dots, q$ are the observations on the gene expression of i and j . This method was used quite often as the result gives a good graphical representation of the hierarchy in the data. In summary two data points in space, defined by two vectors, are compared using Euclidean and Manhattan approaches. A straight line between both points defines their Euclidean distance. If that line is thought to be the hypotenuse of a right angled triangle, the Manhattan distance is the sum of the other two sides.

As there is no "right" or perfect distance function for grouping gene expression values, different methods were tested and applied to find the one showing the best performance on the data set.

All these methods of clustering are so called unsupervised methods, as prior to the analysis there is no knowledge available about the functional roles of the investigated genes. This is not true for all genes but still for the large majority. A hint of the functional role however is given by the analysis process while interpreting the results. Unsupervised microarray data analysis can be seen as an exploratory process where the algorithm discovers the existence of related gene categories without knowledge of the function. The data set is organized based on the properties of data themselves, without any additional information. The clustering methods use a microarray dataset to reveal patterns or groups sharing the same or similar function or a co-regulated. An overview of these clustering methods is given in Figure 38.

The basis of organizing gene expression data in this way is the assumption that similar expression levels might indicate related biological function or co-regulation. The clustering process therefore can be helpful to identify the function of unknown genes (Eisen *et al.*, 1998). During the clustering process expression values are grouped based on a distance function. The genes in one gene expression cluster are similar to other genes in the same cluster, but they are different from genes in other clusters. To perform a cluster analysis the first step is to select an appropriate description of similarity by choosing a distance function. This choice is as important as choosing the clustering algorithm. As there is no universally applicable clustering technique for all different structures present in microarray data the appropriate method was found by testing.

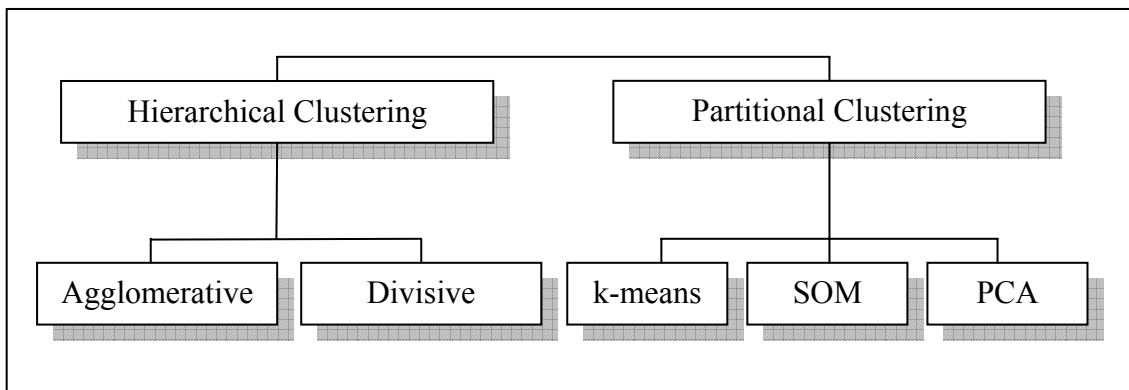


Figure 38: Unsupervised hierarchical clustering methods for microarray analysis. The hierarchical clustering is divided into two groups. The divisive method ("splitting") is based on dissimilarity and the agglomerative ("merging") method based on similarity.

The result of the hierarchical clustering process is a graphical representation as tree like structure called dendrogram. The dendrogram, (Greek word dendron means "tree" and gramma means "drawing"), provides a highly organized visual summary of the clustering process. Such a dendrogram consists of leaves representing gene expression values. The intensity of the colour is a measure of the relative difference between other values and clusters. A red square represents an increase in expression level and a green square shows a decrease. Grey squares represent missing data. A cluster is always connected with two other clusters, two leaf nodes, or one of each. The length of the horizontal lines that connect two nodes of a cluster is a measure of their relative similarity. The hierarchy given by the dendrogram and the visual colour coded representation of the expression levels allows identification of shared expression patterns and therefore closely related genes. A further development called "bi-" or "2D-clustering" is capable of grouping genes as well as experiments (microarrays) simultaneously. This additionally allows identification of correlation between experiments. An example of a 2D-hierarchical clustering is given in Figure 39. It can be seen in this cluster dendrogram that on the experimental level (on top) replicate arrays cluster together. The three arrays "HH26 24 h Rif", "HH26 72 h DMSO" and "HH26 6 h DMSO" and "HH26 6 h Rif" form clusters. In addition the Rifampicin samples (HH26, 27, 44) form a cluster indicating a similar treatment, which is obviously correct. Concerning the gene level CYP3A4 and CYP3A7 build a small cluster. In the same cluster but in a higher distance ALAS1 can be found, indicating a similar expression.

tools are developed which screen publications and data bases for relevant information. One of these tools, Bibliosphere V. 7.05 from Genomatix was applied on 156 genes from the whole genome data obtained from this study. Genomatix BiblioSphere is a data-mining solution to extract and analyse gene relationships from literature databases and genome wide promoter analysis. As an example a correlation search for 156 significant differentially expressed genes from this study was performed. An extract of this analysis is given in Table 27.

Table 27: Spreadsheet view of gene-gene connections obtained from a literature query using bibliosphere version 7.05. The connection between two genes based on co-citations in the literature is given on different levels with an increasing stringency. In the third row the numbers of co-citations in the abstract are displayed. In the fourth row the numbers indicate citations in the same sentence. The fifth row shows co-citation in the same sentence and the sentence also contains a "function word" like "regulation", "inhibit", "transactivate", "target of". The sixth row the two genes are co-cited in one sentence and connected via a "function word" ("gene - function word - gene" connection).

gene 1	gene 2	co-citations on abstract level	co-citations on sentence level	co-citations on function word level	co-citations on gfg level
CYP3A4	ABCB1	282	454	339	87
CYP2C19	CYP3A4	479	564	360	49
CYP2C19	CYP2C9	511	749	426	41
ESR1	IGF1	133	129	103	38
CYP2C9	CYP3A4	541	617	408	30
CYP3A4	CYP3A5	308	512	233	29
CYP2C8	CYP3A4	212	250	148	21
CYP2B6	CYP3A4	245	288	178	12
CYP7A1	HNF4A	31	40	38	12
CYP2A6	CYP2B6	142	164	86	11
CYP3A7	CYP3A4	105	166	72	10
CYP2E1	CYP3A4	300	271	153	9
CYP2A6	CYP3A4	223	219	140	9
CYP2C19	CYP2C8	142	169	93	8
CYP3A4	HNF4A	13	17	17	8
CYP2B6	NR113	38	19	16	8
CYP3A4	NR113	36	19	16	8
IGF1	SOX9	10	10	9	7

In the example shown in Table 27 CYP3A4 and ABCB1 are 282 times in the same abstract of any publication, they appear 454 times in the same sentence in the scientific literature. In 339 times the sentence in which they are used together contains in addition a functional word like activate, regulate, repress, stimulate and induce. In 87 of these cases both genes are connected with such a "functional word". Further you can find that

CYP3A4 has a strong correlation with a lot of other CYPs as well as with HNF4A and NR1I3 (CAR). Furthermore information for all Pubmed abstracts compiled into this “BiblioSphere” can be displayed as shown in Table 28.

Table 28: Overview of pubmed listed articles containing genes which were found in this study to be significant differentially expressed. Given is the access to the literature (PMID), the genes which are mentioned in this literature and the overlap number of genes important in the study presented here and given in the respective publication.

PMID	Genes	# Genes
17327493	CITED2, ELL2, MAFF, EGR2, FOS, FOSB, NR4A1, MITF, NFIL3, NR4A2, PER1, SALL1, NR4A3, BHLHB2	14
15558475	CREB1, HSF2, JUN, MITF, NRL, PAX6, PROX1, RARB, RORA, SIX3, SOX1, SOX2, TBP, MAFB	14
16027002	CITED2, HES6, GATA2, ID1, LMO1, MYB, MYBL1, MYC, MYCL1, MYCN, NHLH1, SOX1, TEAD2	13
9242638	ID1, ID2, ID3, LYL1, MYF5, MYF6, MYOD1, MYOG, TAL1, TAL2, TCF4, TCF12	12
12815626	SOX30, SOX8, SOX18, SOX17, SOX7, SOX4, SOX5, SOX9, SOX10, SOX11, SOX15, SOX13	12
14634838	ADH1A, ADH1B, CYP2C19, CYP2C9, CYP2E1, AHR, ESR1, ESR2, NR3C1, HSD17B2, PGR	11
15858847	AHR, NR5A2, NR3C1, HNF4A, PPARA, TCF1, VDR, NR0B2, NR1I2, NR1I3, NR1H4	11
15642787	CEBPA, CEBPB, DBP, FOXA1, FOXA2, FOXA3, HNF4A, ONECUT1, STAT5A, TCF1, TCF2	11
15499180	CYP3A4, NR2E3, CYP3A5, ESR1, NR5A2, NR5A1, HNF4G, THRA, NR2E1, NR1I3, NR1H4	11
11230748	CYP2A6, CYP2B6, CYP2C9, CYP2E1, CYP3A4, CYP3A5, HNF4A, RXRA	8
15698577	CYP3A7, CYP3A4, CYP3A5, NR3C1, HNF4A, CYP3A43, NR1I2, NR1I3	8
9893992	DR1, NR2F6, HNF4A, PPARA, RARA, RXRA, SP1, NR2F1	8

This investigation allows the researcher to identify very fast and easily the most important and interesting publications for his data set. In this example here there are several publications showing the connection between CYPs (the metabolising enzymes) and HNF4 α or NR1I3 (nuclear receptors).

In this small section presented in Table 28 several publications were identified containing more than ten of the relevant genes. The first article e.g. (PMID: 17327493 [Pubmed - indexed for MEDLINE]) “Adrenal transcription regulatory genes modulated by angiotensin II and their role in steroidogenesis.” reports a set of transcription regulatory genes that are modulated by angiotensin II and their role in adrenal gland steroidogenesis. As several genes are also differentially expressed in this study, they

might be also altered by Rifampicin. Abnormal regulation of the mineralocorticoid or glucocorticoid biosynthesis pathways is involved in several pathophysiological conditions. The modulated transcription regulatory genes described in this paper and found in our study may correlate with pathologies in adrenal steroidogenesis.

Finally this data can be visualized in a three dimensional (3-D) space. The 3-D view reflects the co-citation frequency as the distance between two genes. The more often two genes occur in the same abstract or sentence, the closer they are in the "Literature Molecule". The 3-D graphs (BiblioSpheres) allow the identification of complex gene relation structures. You can navigate the literature and get information on how often two genes are co-cited by moving over the connecting lines between the spheres that represent these genes. Additional information on focused genes and their relationships can be also obtained. If one of these genes is a transcription factor (TF), additional information on potential binding sites in the promoters of its co-cited neighbours are displayed. Connections affirmed by this promoter analysis are displayed in green. Figure 40 shows the interconnection of co-cited genes to CYP3A4.

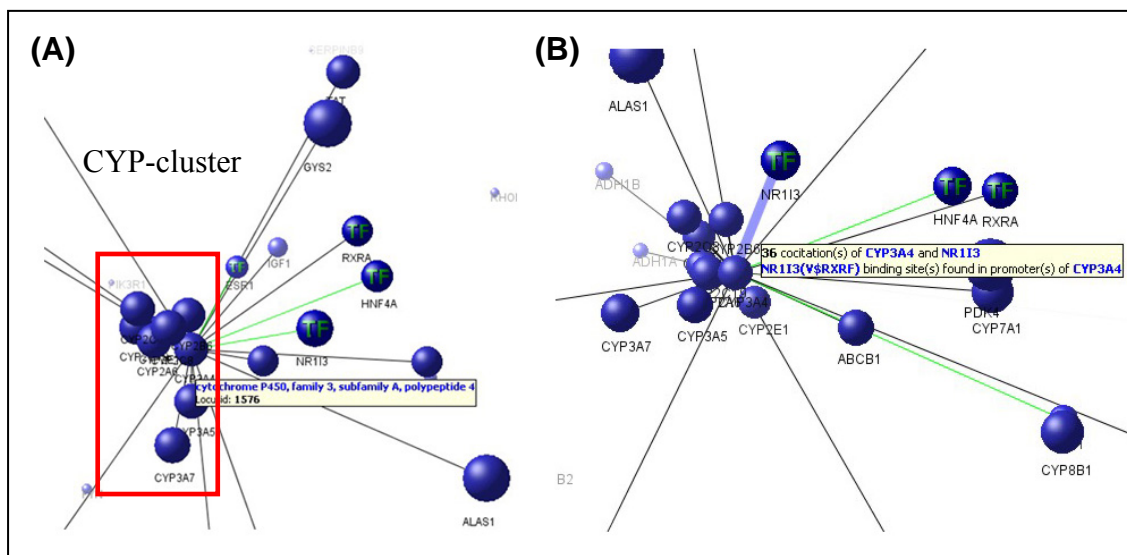


Figure 40: Three dimensional visualization of co-citation of genes in scientific literature ("BiblioSphere"). Every gene is represented by a blue sphere. The distance between spheres represents their relationship to each other based on the literature. (A) Shows the interconnection of co-cited genes to CYP3A4. The literature based proximity is obvious. In (B) the relationship to the three transcription factors (TF) NR113, HNF4A and RXRA can be clearly seen. As they are shown in green this relationships are confirmed by scientists. As displayed in the graph 36 co-citations were found between CYP3A4 and NR113.

As previously shown in the spreadsheet view (Table 27) this CYP has a tight literature based interconnection to other members of the CYP family. This can be easily seen from the three dimensional plot, as the distance is a measure of connectivity. Several members of the CYP family form a cluster (indicated with a red rectangle) showing a close relationship. In addition the connection to ALAS1 and ABCB1 is shown. Concerning regulation it can be seen that there is connection to three transcription factors (NR1I3, RXR α and HNF α). For CAR (NR1I3) for example 36 co-citations with CYP3A4 were identified (as indicated in Figure 40 B).

3 Results

The use of microarray based analysis for high throughput screening of differentially expressed genes in primary hepatocyte cell culture upon stimulation with drugs was evaluated in this study. Two different approaches were used to collect expression data for a systems biology approach to model the response of metabolic and regulatory networks in the human liver.

On one hand a so called "top-down" approach by using a full genome DNA chip from Affymetrix. This procedure allowed by measuring the expression of all human genes a holistic functional analysis of the hepatocyte in response to the perturbation. Based on this data set the reconstruction of metabolic, regulatory and signal transduction networks was performed using a literature mining and microarray analysing (LMMA) approach.

On the other hand based on already available biological and medical knowledge a sub-genome array covering 151 genes involved in liver metabolism was used to extend the findings in "bottom-up" experiments. This inductive procedure investigates single parts and components of the whole system. Using small density arrays has two advantages: due to the smaller number of investigated, well annotated genes, they are less expensive and easier to interpret.

Primary human hepatocytes were obtained from patients undergoing a liver surgery. The cells were cultivated and induced in six well plates. This experimental procedure was used to analyse the expression of genes using microarrays after stimulating the cells with Rifampicin. Therefore Rifampicin diluted in DMSO was added to one part of the cell culture while a second part of the same culture was incubated parallel with DMSO only. After 0 hours - time of induction - 6, 24, 72 hours samples were taken for the full genome analysis. A subset of liver specific genes was monitored after 6, 12, 16, 20, 24, 28 and 72 h using the low density arrays. Expression ratios from treated and control cultures were determined. To investigate the induction of the cell cultures, quantitative real-time PCR (qRT-PCR) was performed to verify the up-regulation of some already known key-players of the detoxification machinery (CYP3A4, CYP2D6).

In case of successful induction, total RNA was reversely transcribed or amplified and labeled. After determining the quality of amplified labeled RNA, the samples were hybridized on the arrays. For detection of the fluorescence signal, the arrays were scanned and the signals were quantified. Finally the raw data was processed to obtain relative expression ratios for the differences in gene expression of induced versus not induced cells. Based on the obtained expression levels, further data analysis like clustering and network-reconstruction was done to gain insight in the responding cellular functions. The data set was also the basis for a modelling approach in cooperation with the Institute of Biochemical Engineering. Since a quite tight sequence of sampling is necessary for dynamic modelling, which is hardly affordable using DNA-microarrays due to cost constraints, a qRT-PCR based method was used. The change of mRNA levels for CYP3A4 and PXR were detected quite frequently during a large time series experiment. CYP3A4 is the predominant metabolic enzyme detoxifying Rifampicin and PXR is one of the most important nuclear receptors regulating this detoxification process in general. In this investigation samples were analysed after 0.5, 1, 4, 8, 12, 16, 24, 30, 36, 42 and 48 hours.

3.1 Comparison of Microarrays from Different Suppliers

Before analysing the scientific question regarding the response of human hepatocytes to Rifampicin, a comparison of several microarrays from different suppliers and general considerations concerning the microarray technology were done to investigate whether the approach using different kind of platforms and principles would result in useful and comparable results.

After a comprehensive literature screening three commercial microarray platforms were chosen to be evaluated in the laboratory. The selected full genome array systems evaluated in the laboratory were from Agilent and Affymetrix. In addition the sub-genome liver specific (Hepato-) DualChip from Eppendorf was compared. Figure 41 summarizes the features of the selected array systems.



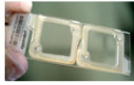
	Affymetrix	Agilent	Eppendorf
			
Number of transcripts	> 44.000	> 44.000	151
format	special	standard glas slide	standard glas slide
starting material	1 – 2 µg *	1 – 2 µg *	5 – 10 µg 1 – 2 µg *
data evaluation	sophisticated	sophisticated	simple
costs for one experiment	2200 € #	875 € +	160 € +

Figure 41: Summary of features for different microarrays used in the comparison study of this work. The total number of transcripts, the hardware of the array, the needed amount of mRNA per experiment, the kind of data analysis and the cost for analyzing two samples is given. * 1 - 2 µg are sufficient if an amplification step of RNA is performed which is always done for the full genome arrays. The Eppendorf array also allows detection of not amplified samples. # for two arrays, as Affymetrix is a one colour platform two arrays are required to analyse a control versus an experimental sample. + Agilent is a two colour platform so both samples can be hybridized on one array and Eppendorf allows both one-colour as well as two-colour labeling, therefore only one slide is necessary to determine the ratio between two samples (prices Nov. 2005).

The most important differences between Affymetrix and Agilent arrays are the hardware-format and the price. As Affymetrix uses its own hardware the experimental work has to be contracted to a microarray facility as the hybridization station and microarray scanner have an Affymetrix specific format. The Agilent arrays are based on standard glass slides, so it was considered to hybridize and scan these arrays in house. However this was not possible as the high number of spots of a human full genome array required a special format of the grid (arrangement of the spots) which was a sphere packing like structure which the Perkin-Elmer scanner (ScanArray Express) used at the Institute of Technical Biochemistry was not able to resolve.

The second major difference is the price. Both arrays cost about 1000 € however for investigation of the ratio between two samples the Affymetrix platform requires two one colour chips to measure the expression in sample 1 and in sample 2. The Agilent array is used as two colour array which allows hybridization and detection of both

different labeled samples on one array. Therefore only one array is needed (as long as no dye-swap experiment is performed) yielding to only half of the costs. As Affymetrix is still the "gold standard" in the field of DNA microarrays, it is further well established in the community, meaning a lot of analysis tools are available and the acceptance of results is quite high, Affymetrix chips were preferred for the full genome analysis in the work presented here.

Although the information content using the Eppendorf array is much lower, it was an interesting alternative, as it costs less than 1/5 of the price of a full genome chip. This kind of low density array was used to investigate the expression level of the 151 available genes in more samples from a time series experiment collected with a higher sampling frequency.

The overall intra platform reproducibility was very high. Correlation of expression ratios measured on individual arrays of one cell culture (HH26) was determined to be 0.93. The Eppendorf platform provided also highly reproducible (0.8) measurements of gene expression profiles, which was an essential pre-requisite for the cross-platform comparison.

Three array platforms for gene expression profiling from Affymetrix, Agilent and Eppendorf were compared in our laboratory to analyse their performance with RNA samples obtained from primary hepatocyte cell cultures to decide which platform was the most suitable for our purpose.

On all three different platforms the same samples were analysed based on the information from the supplier. Induced as well as not induced control samples were hybridized on the arrays and the expression levels of several genes known to be involved in the detoxification were compared. Furthermore the results for two CYPs were confirmed using real time PCR and the induction of CYP3A4 was also validated by western-blot. The results of this comparison are given in Table 29.

Table 29: Expression values for relevant detoxifying enzymes, obtained using three different microarray platforms. The expression values of CYP2B6 and CYP3A4 were validated using Taq-Man assays and the induction of CYP3A4 was also validated on protein level by Western-blot.

Gene	Affymetrix	Agilent	Eppendorf	Taq-Man	Western blot
CYP 1A2	1.0	1.0	0.8		
CYP 2A6	1.4	1.6	0.6		
CYP 2B6	4.0	2.8	2.1	5.9	
CYP 2D6	0.8	0.8	0.6		
CYP 3A4	2.4	3.7	2.5	2.9	2 - 3 fold
MDR1	3.2	1.4	3.8		
MRP2	1.7	2.1	1.3		
MRP3	1.0	0.8	1.0		
GSTA1	2.3	2.1	2.5		
EPHX1	2.9	2.3	2.6		
UGT1A6	2.0	2.0	2.2		
PXR	0.5	0.6	0.8		
SULT 1A1	0.8	0.9	0.8		

For the genes investigated in this study, which are important in detoxification of xenobiotics and therefore are main targets for the modelling approaches, a good correlation between the different array platforms was observed. The comparison of array results with results obtained by Taq-Man (qRT-PCR) and Western-blot were also satisfying. The induction of CYP2B6 was between 2.1 and 4.0 based on the array measurements and monitored to be 5.9 fold induced using a Taq-Man based assay. CYP3A4 was measured to be 2.4 to 3.7 fold induced by microarray analysis and 2.9 by a Taq-Man based assay. Western-blot detection revealed a 2 to 3 fold increase of CYP3A4 protein level. As there were almost no differences concerning handling and performing of sample preparation for Agilent-arrays and the Affymetrix platform the more expensive, but in the research community and market well established Gene Chips from Affymetrix were chosen as full genome platform. A comparison of the results obtained with different platforms is given in Figure 43.

Figure 42 shows a correlation plot between the different applied methods to determine the expression of genes in human hepatocyte cell cultures is given. The fold of induction determined with Affymetrix chips is plotted over the two other microarray platforms from Agilent and Eppendorf as well as two genes determined with real-time PCR (TaqMan). Between the array platforms a correlation of 0.6 was determined for a subset of 20 genes involved in the xenobiotic metabolism. The gene list overlap and

concordance of differentially expressed genes was good. The ratio or fold-change analysis yielded in unequal results based on different approaches to measure the expression and determine the fold change. It was observed that e.g. the TaqMan fold-change values were higher compared to results obtained from the Affymetrix arrays. The general overlap of detected genes was high but comparable quantitative results between different platforms could only be obtained for a small subset of genes.

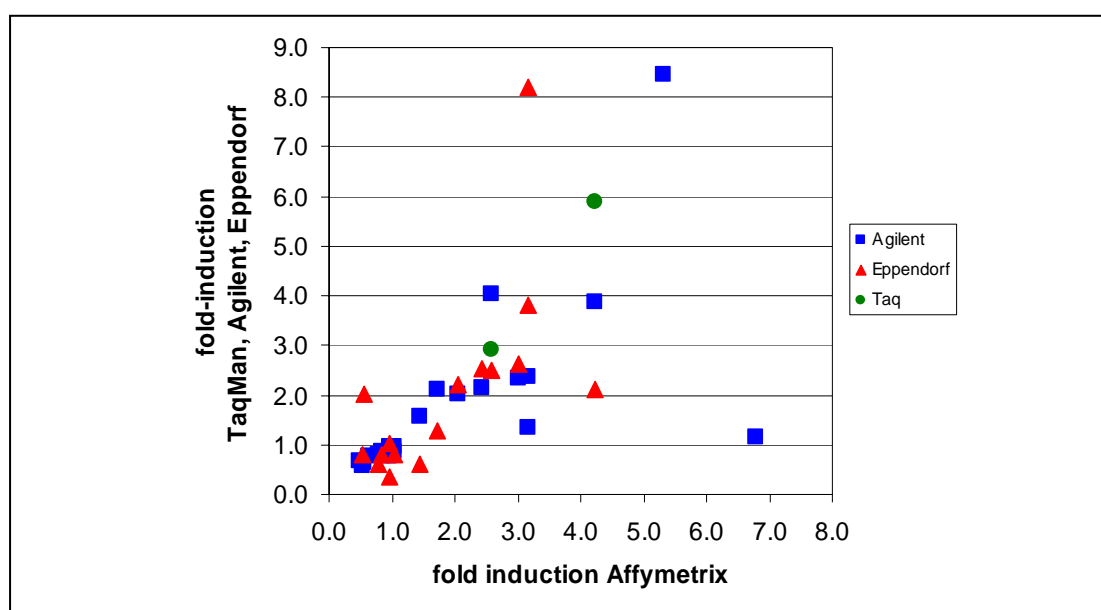


Figure 42: Correlation between Affymetrix (n = 22) and TaqMan (n = 2), Agilent (n = 21) and Eppendorf (n = 17) data. The scatter plot compares the fold changes of differential expression values in human hepatocyte culture HH22 after 24 h Rifampicin induction. The correlation for between Affymetrix and Agilent was 0.62. The comparison of Affymetrix and Eppendorf data gave a similar correlation of 0.63.

Further comparison between Affymetrix and Eppendorf was performed. The results of similar detected genes on both platforms from the same samples are given in Figure 43.

A good correlation between the two platforms for this set of genes was observed. In the next step, a whole experiment with three different points in time (6, 24, 72 h) was performed in parallel with both systems. The genes showing a similar gene expression ratio in both experimental setups are given in Figure 44 - Figure 46.

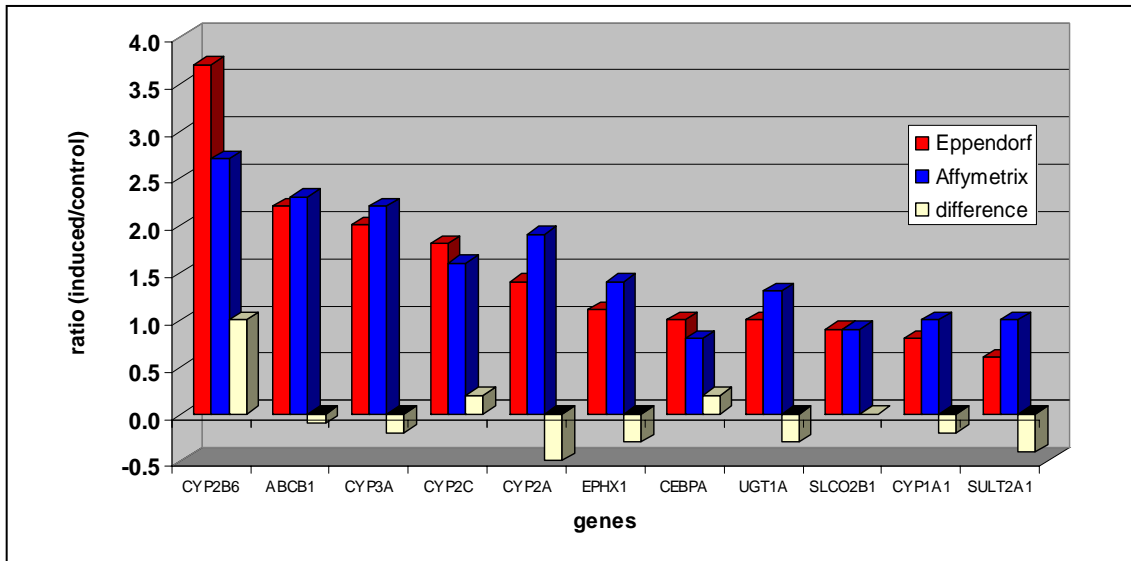


Figure 43: Comparison of expression levels obtained from Eppendorf and Affymetrix with primary hepatocyte cell culture HH26. Given in red are the expression levels detected with Eppendorf Arrays. Shown in blue are the expression levels monitored with Affymetrix chips from the same samples. White bars indicate the differences between both measurements (positive for a higher value from the Eppendorf data, negative if the Affymetrix chips detected a higher signal).

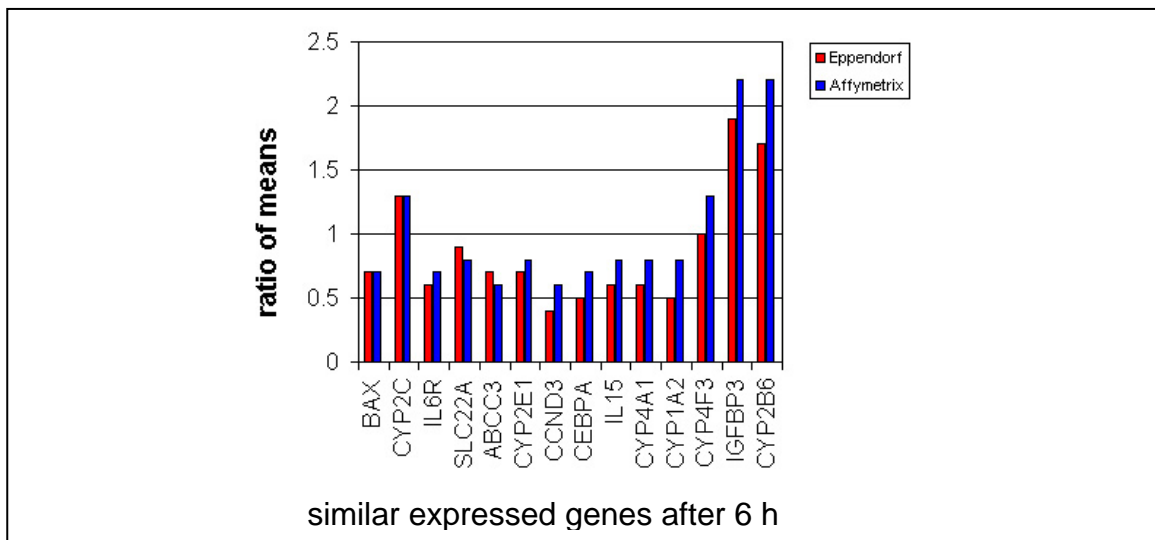


Figure 44: Genes showing a similar gene expression ratio after six hours of induction with Rifampicin on Eppendorf DualChip arrays and Affymetrix full genome GeneChips. Ratio of means between Rifampicin treated cells versus control (DMSO) cells. Levels measured with Eppendorf Chips are given in red and Affymetrix data is shown as blue bars.

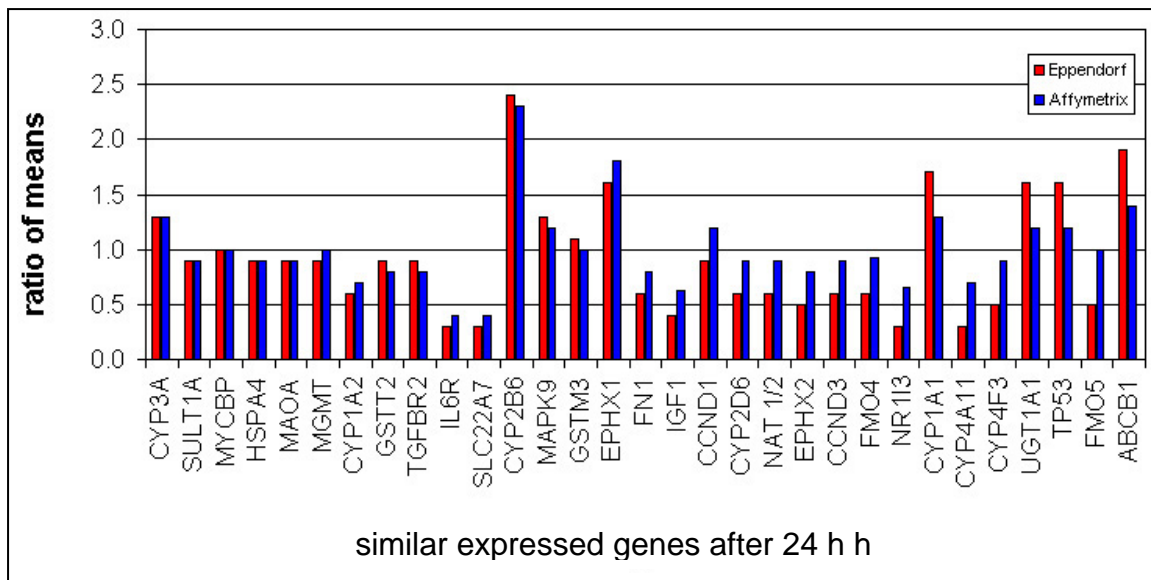


Figure 45: Genes showing a similar gene expression ratio after 24 hours of induction with Rifampicin on Eppendorf DualChip arrays and Affymetrix full genome GeneChips. Ratio of means between Rifampicin treated cells versus control (DMSO) cells. Levels measured with Eppendorf Chips are given in red and Affymetrix data is shown as blue bars.

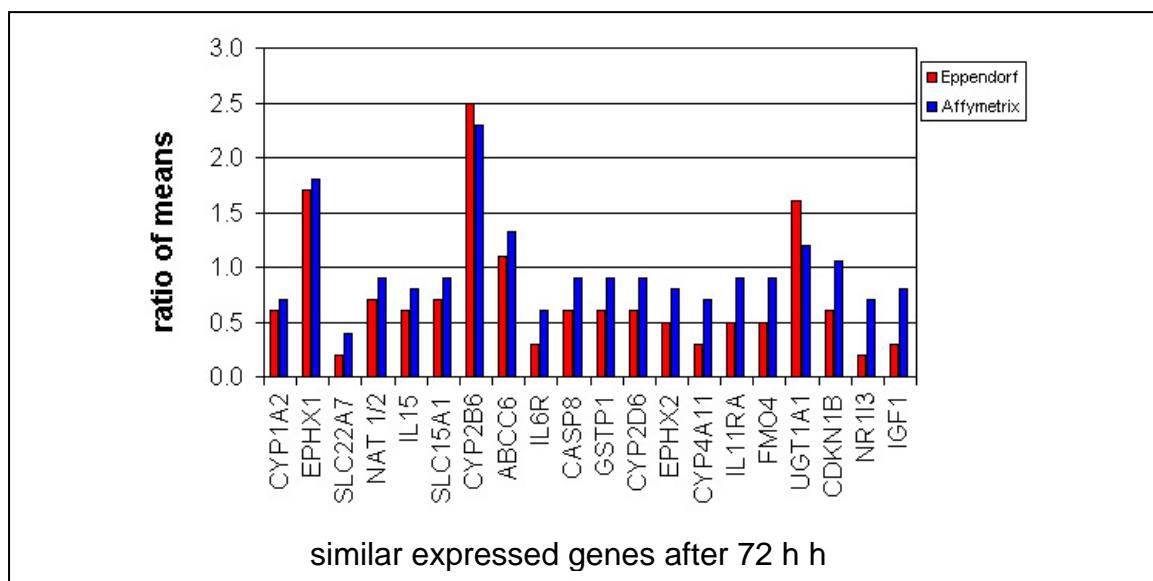


Figure 46: Genes showing a similar gene expression ratio after 72 hours of induction with Rifampicin on Eppendorf DualChip arrays and Affymetrix full genome GeneChips. Ratio of means between Rifampicin treated cells versus control (DMSO) cells. Levels measured with Eppendorf Chips are given in red and Affymetrix data is shown as blue bars.

These experiments prove that by both array systems the up-regulation of CYP2B6 for all three points in time could be measured. Both platforms show similarly, that IL6R, SLC22A7 and CYP4A11 were down-regulated. After 6 hours of induction a strong up-

regulation of IGFBP3 and after 24 hours of CYP3A4 was detected by both platforms. After 24 h and 72 hours EPHX1 and UGT1A1 were up-regulated while NR1I3 and IGF1 were monitored to be down-regulated.

Although there was a good overlap of 14, 31 and 20 genes (in the different measurement points), the overall correlation between the array platforms for all genes was, as previously reported (e.g. in "The MicroArray Quality Control MAQC project" Shi *et al.*, 2006), poor. However each array platform delivered consistent and reproducible data, but the comparison between the platforms was limited to a small subset of genes.

The genes showing similar expression levels for the different platforms were: cytochrome P450, family 1, subfamily A, polypeptide 2 (CYP1A2), cytochrome P450, family 2, subfamily A, polypeptide 6 (CYP2A6), cytochrome P450, family 2, subfamily B, polypeptide 6 (CYP2B6), cytochrome P450, family 2, subfamily C, polypeptide 9 (CYP2C), cytochrome P450, family 2, subfamily D, polypeptide 6 (CYP2D6), cytochrome P450, family 3, subfamily A, polypeptide 4 (CYP3A4), cytochrome P450, family 4, subfamily A, polypeptide 11 (CYP4A11), interleukin 6 receptor (IL6R), insulin-like growth factor binding protein 3 (IGFBP3), solute carrier family 22 (organic anion transporter), member 7 (SLC22A7), epoxide hydrolase 1, microsomal EPHX1, UDP glucuronosyltransferase 1 family, polypeptide A1 (UGT1A1), UDP glucuronosyltransferase 1 family, polypeptide A6 (UGT1A6), nuclear receptor subfamily 1, group I, member 3 (CAR NR1I3), nuclear receptor subfamily 1, group I, member 2 (PXR NR1I2), insulin-like growth factor 1 (IGF1), ATP-binding cassette, sub-family C (CFTR/MRP), member 2 (ABCC2, MRP2), ATP-binding cassette, sub-family C (CFTR/MRP), member 3 (ABCC3, MRP3), glutathione S-transferase A1 (GSTA1) and sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 (SULT1A1).

For the Affymetrix platform it had to be taken into consideration that each gene is represented by different probe sets on one array. These probe sets are based on different complementary sequences of the same target gene. For this target gene eleven perfect match and eleven miss match signals are measured and then during the condensing algorithm merged to one signal value which is exported for a set of probe pairs. The different sets of probe pairs however are not combined together. Different values for

different probe pairs of the same target gene are obtained. This makes the comparison with other platforms more difficult.

As an example the expression ratio of CYP3A4 detected by two different probe sets on the Affymetrix array are plotted in Figure 47. It can be seen that for the same gene in the same sample, on the same array different fold changes were obtained. The induction varied from 6-fold induction after 24 h for probe-set 205998_x_at to only 2.5-fold induction for probe 205999_x_at. After 72 h a similar result was observed. The induction of CYP3A4 was 8-fold based on probe 205998_x_at and 4-fold for the same gene based on probe 205999_x_at.

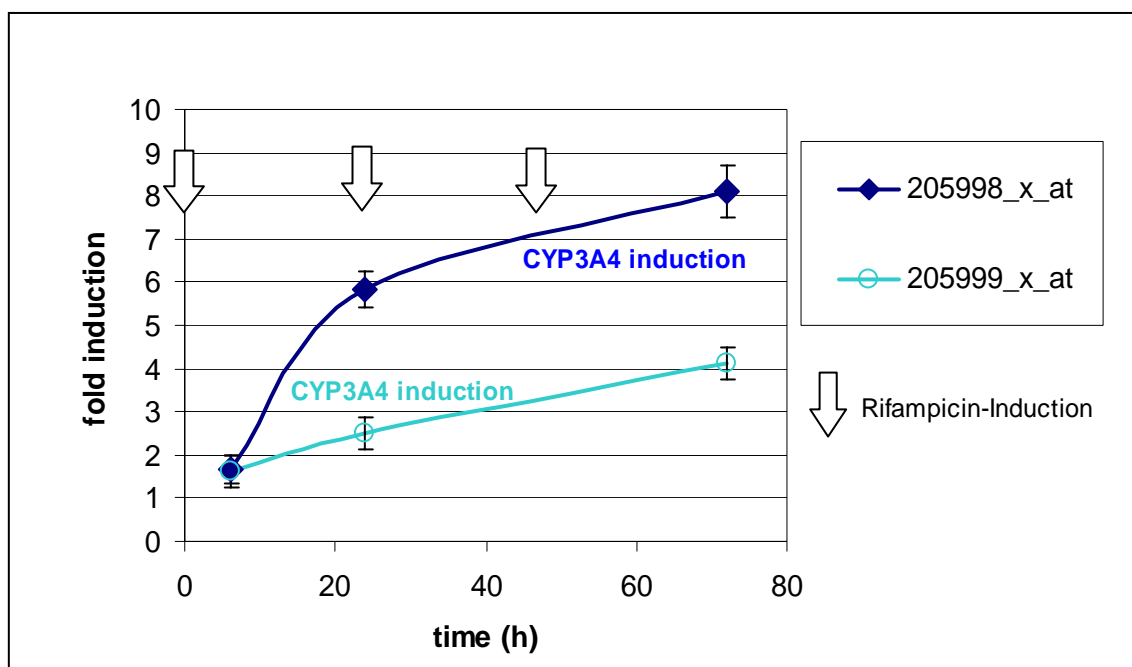


Figure 47: Expression ratio of CYP3A4 on the same array from the same sample with different probe sets. Based on probe set 1 (205998_x_at) the expression of CYP3A4 was 6 fold after 24 hours and up to 8 fold after 72 hours. If probe set 2 (205999_x_at) on the same array is used, a 2.5 fold increase after 24 hours and a maximal increase of 4 after 72 hours is obtained.

3.2 Experimental Setup of Time-Series Experiments

The cultivation of primary human hepatocyte cells was done in cooperation with the Institute of Biochemical Engineering, University of Stuttgart.

For experiments with three measurement points (6, 24, 72 h) and two treatments (Rifampicin and DMSO as control) six six-well plates were used resulting in 36 wells. At every point in time and for each treatment, cells were harvested in triplicates (3 wells). In case of loss of several wells, due to insufficient attachment of the cells to the surface or contamination, the number of replicates was reduced.

For larger time series experiments different experimental setups were prepared as the number of available plates often varied and in some cases had to be changed due to loss of attachment or contamination during transport. Therefore the suitable setup was chosen one day after arrival of the cells based on the amount of usable cells. An overview of different setups is given in Figure 48.

Expression profiling experiments performed, using the previously mentioned cell culture conditions and samples are summarized in the following overview given in Figure 49. Light blue rectangles in this graph represent measurement points for taking samples which were analysed using Affymetrix full genome arrays. For cell cultures from patient HH26 three replicate measurements from the same cell culture but different preparations were performed. This analysis of technical replicates allowed the estimation of technical variance of our setup. To analyse the more interesting and important biological variance two more replicates (HH27 and HH44) were investigated under the same conditions and for the same points in time.

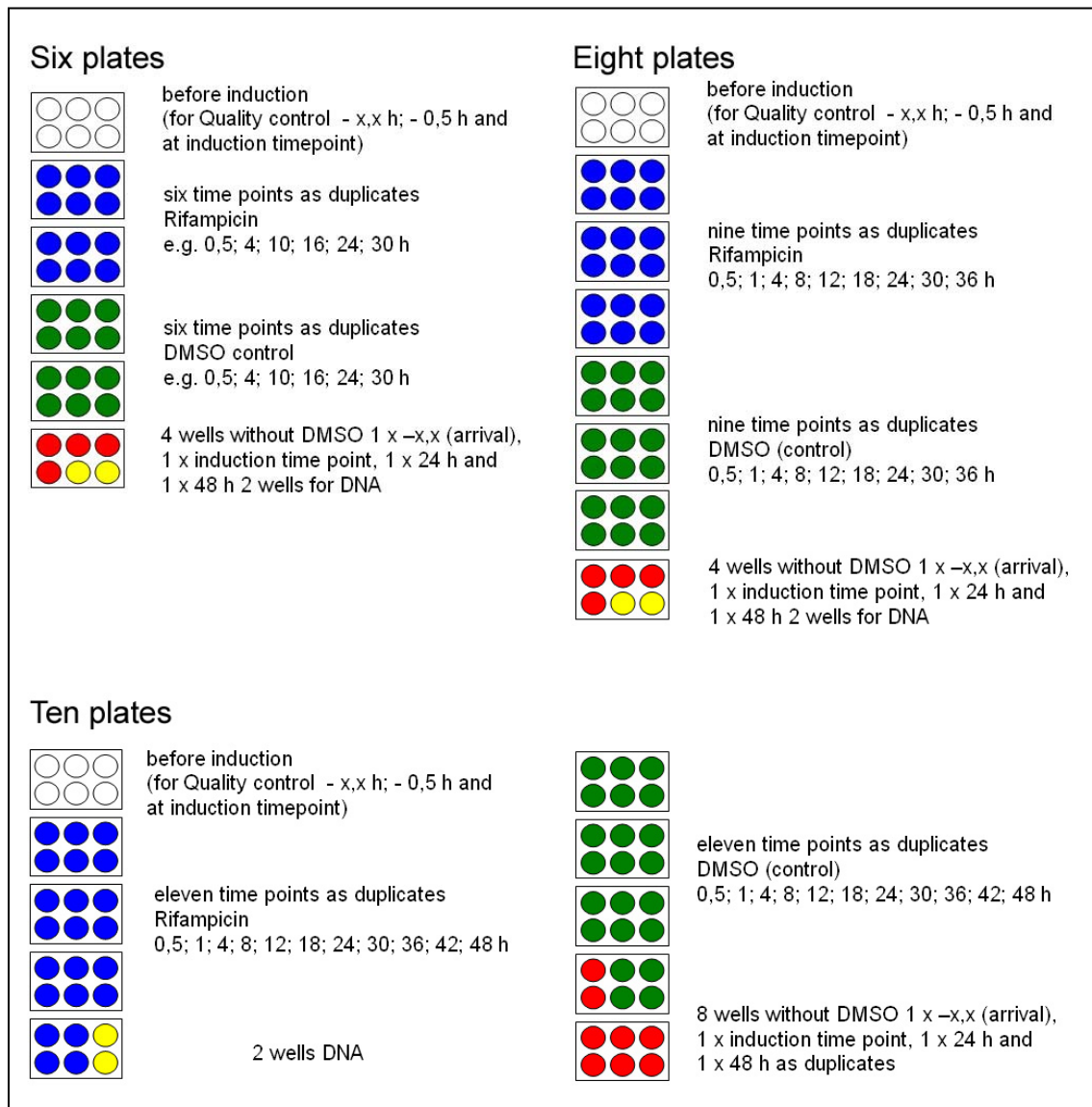


Figure 48: Overview of experimental time series setups for different amounts of delivered cells. Plate allocation for hepatocyte cells in six well plates is indicated by color. Blue stands for Rifampicin treated cells, green for control cells (DMSO), red for untreated cells and yellow for wells from which DNA was isolated. White plates were used for determining the quality of the hepatocyte cell cultures. All RNA samples (DMSO, Rif, untreated) were done in replicates. Samples without any treatment were taken four times (at arrival of the cells, at the induction point and after 48 h) to investigate dedifferentiation of cells in culture and the impact of DMSO on the cells.

To enhance the assessed full genome information, two more measurements with an increased number of measured points were performed. For this purpose sub-genome arrays from Eppendorf (light blue circles) were used. Sample HH38 was induced in the same way as the samples for the full genome Affymetrix array (shown by red arrows) and samples were taken after 2, 8, 16, 24, 32 and 63 h to investigate the dynamic change in mRNA concentrations during this induction process. A third approach was

performed, using Eppendorf DualChips and sample HH48, to get information about the regulatory system by removing the inducing agent (Rifampicin, indicated by red arrows in opposite direction). A high sampling frequency (4 h) around this perturbation was measured. First the induction was analysed and proven by measuring samples after 6, 12, 16, 20 and 24 h. After 24 h the media was changed to media without inducer. 4 h after this media change (28 h) the short term effect was measured on an array. After 72 h another sample was taken and analysed to see whether the mRNA levels were back to their starting concentration or still increased.

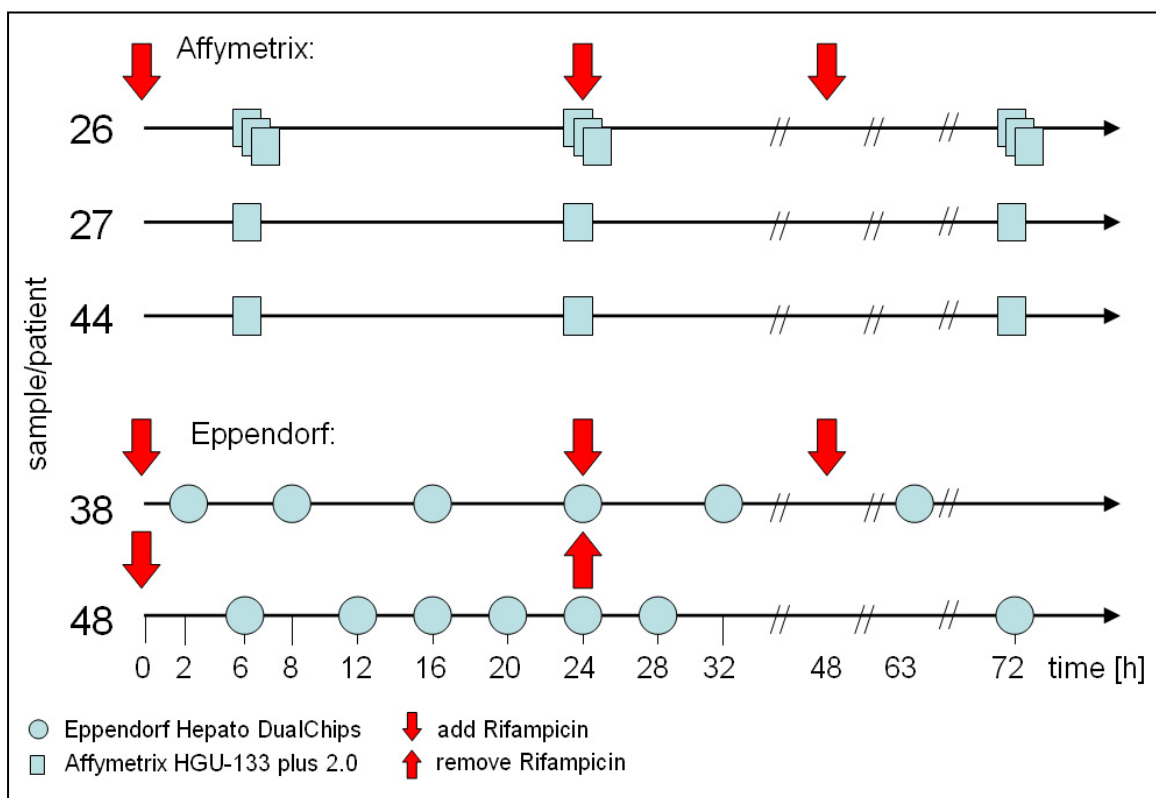


Figure 49: Experimental setup for the investigation of transcriptional response of primary human hepatocyte cells upon Rifampicin induction. The numbers on the left (26, 27, 38, 44 and 48) indicate the samples from different patients used for this study. The circles and boxes show the sample points when mRNA was extracted and analysed on the full-genome Affymetrix and sub-genome Eppendorf arrays. The arrows mark the points of induction and removal of Rifampicin.

3.3 RNA Isolation and Quality Control

After cultivation, induction and harvesting of hepatocyte cells, the RNA had to be extracted. Quality and quantity are important issues especially when working with a limited amount of primary cells. To obtain optimal results several RNA extraction methods were compared. For higher reproducibility and convenience the phenol chloroform extraction which leads also to high quality RNA was also tested, but in the end not used for routine RNA extraction. Finally kits based on organic extractions, cell lysis in presence of chaotropic agents and adsorption of the RNA on solid supports (silica based spin columns) from three suppliers were compared. All three kits from Qiagen, Agilent and MacheryNagel, tested in this study yielded very high quality RNA (Figure 50).

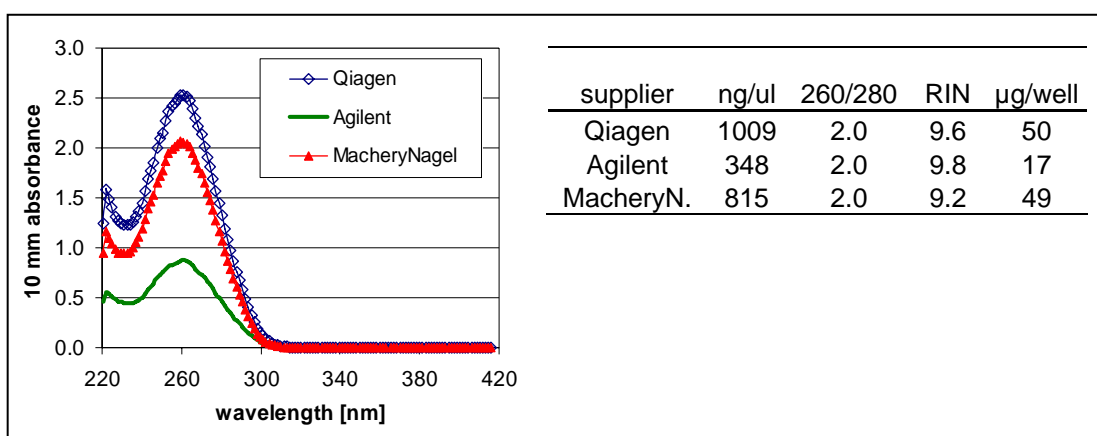


Figure 50: Comparison of RNA isolation kits from different suppliers. Given is the absorption profile of isolated RNA samples (left), the concentration obtained, 260/280 absorption value, the RIN and the amount of RNA/well (1.5 mio hepatocyte cells).

Concerning RNA quantity in the comparison experiment, the Qiagen and MacheryNagel kits yielded almost the same amount of 50 µg RNA/well, while the amount of RNA recovered with the Agilent kit was much lower in this direct comparison. However in subsequent tests the Agilent kit showed a similar performance as the Qiagen and MacheryNagel kits. Based on quality and quantity of RNA, price and convenience, the Qiagen RNA extraction kit was selected. A further investigation was if Trizol (guanidinium thiocyanatephenol, RNase inhibitor) has an impact on the quality of RNA and whether freezing leads to degradation. The results are summarized

in Figure 51. In all cases high amounts of RNA were obtained and almost no degradation could be observed in any case. Finally it was concluded that for our experiments the use of Trizol is not necessary and that freezing of cell suspension in RLT-buffer (from the Qiagen kit see 2.7.3) does not affect the quality of isolated RNA. The concentration and purity could be further increased by ethanol precipitation. However it has to be taken care, to avoid degradation and carry over of ethanol. Residual ethanol might interfere with subsequent reactions.

Extraction of total RNA was done based on the Qiagen Mini kit protocol as described in the material and methods part (2.7.3). The RNA was isolated, frozen in liquid nitrogen and stored in aliquots for measuring the abundance of the different transcripts by real-time PCR and microarray analysis.

To assess the concentration of RNA samples UV/Vis spectra were measured using a NanoDrop spectrophotometer. The concentration of RNA, given in Figure 50 is the amount isolated from one well of a six well plate (about 1.5 million hepatocyte cells). To determine the degradation status of RNA (RIN = RNA integrity number quantification) capillary gel electrophoresis (Agilent Bioanalyzer 2100) was performed. Alternatively to capillary gel electrophoresis a standard agarose/formaldehyde gel electrophoresis was performed as shown in Figure 51.

As shown in Figure 52 the concentration for the 6 h point was a bit lower compared to 24 and 72 h and ranged from 89 to 290 ng/ μ l. The amount of RNA isolated after 24 h was between 277 and 437 ng/ μ l and for the 72 h samples it ranged between 231 and 510 ng/ μ l.

The 260/280 ratio was for all samples 2.0 suggesting a high purity of isolated RNA. The 260/230 ratio showed in some cases a lower value indicating some impurities which can be also seen in the graphical display of the spectra in Figure 52.

The absorption curve which has its maximum at 260 nm, for the absorption of nucleic acids, show in some cases another high value at low wavelength (230 nm) leading to a low 260/230 ratio. These contaminations occurred in some samples like "9 d3 6h DMSO 1" or "3 d3 6h Rif 3" resulting in a 260/230 ratio <1.0. However these impurities did not affect the following amplification and labeling reactions. Samples with low 260/230 ratios showed a similar behaviour to samples having a high 260/230 value.

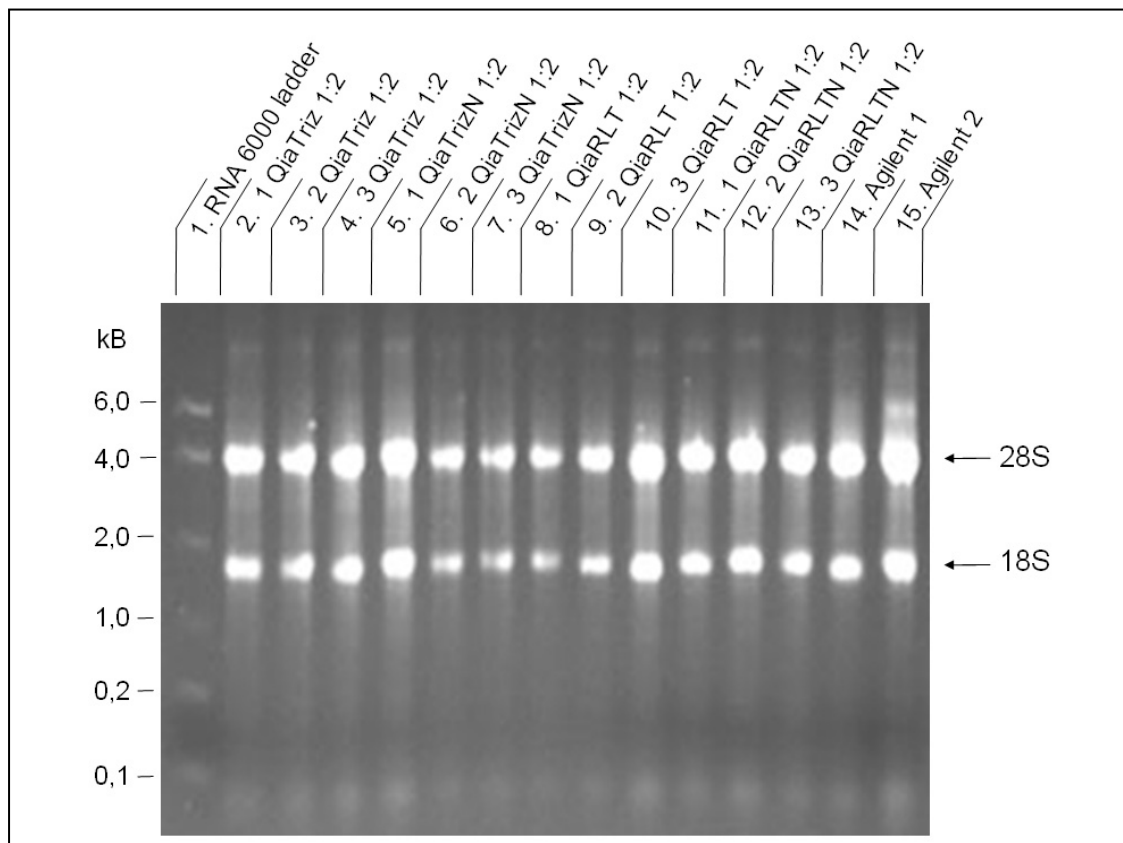


Figure 51: 1.2% ethidium bromide stained agarose/formaldehyde gel showing the comparison of different RNA extraction methods based on Qiagen and Agilent RNA extraction kits. 2 μ l of a 1:2 dilution (except for Agilent samples which are undiluted) were separated for 40 min at 100 V = const. and 50 mA. Lane 1 shows the RNA size standard from Ambion (RNA 6000). The lanes 2 till 15 show RNA extracted with different methods. In all cases the two ribosomal RNA bands (28S- and 18S-RNA) are visible and not degraded (arrows). Lanes 2, 3 and 4 show RNA, which was stored in Trizol, precipitated in EtOH and extracted with the Qiagen Mini Kit. Lanes 5, 6 and 7 show the same extraction method as the three previous ones, except that the trizol treated cells were frozen in liquid nitrogen. Lanes 8, 9 and 10 show the RNA which was isolated using the standard protocol for the Qiagen Mini Kit (RLT-Puffer for cell lysis). RNA in lanes 11, 12 and 13 was extracted with the standard protocol (Qiagen Mini Kit) but before purification frozen in RLT buffer with liquid nitrogen. Lanes 14 and 15 show RNA which was isolated using the Agilent RNA-isolation kit (based on the protocol of the supplier). More or less all tested methods yielded good quality RNA with almost no degradation.

A complete overview of all samples (from culture HH26) characterised by absorption measurement is given in Figure 52. 17 RNA samples showed a lower 260/230 ratio compared to 260/280 which was for all samples in the expected range. All samples were of high purity.

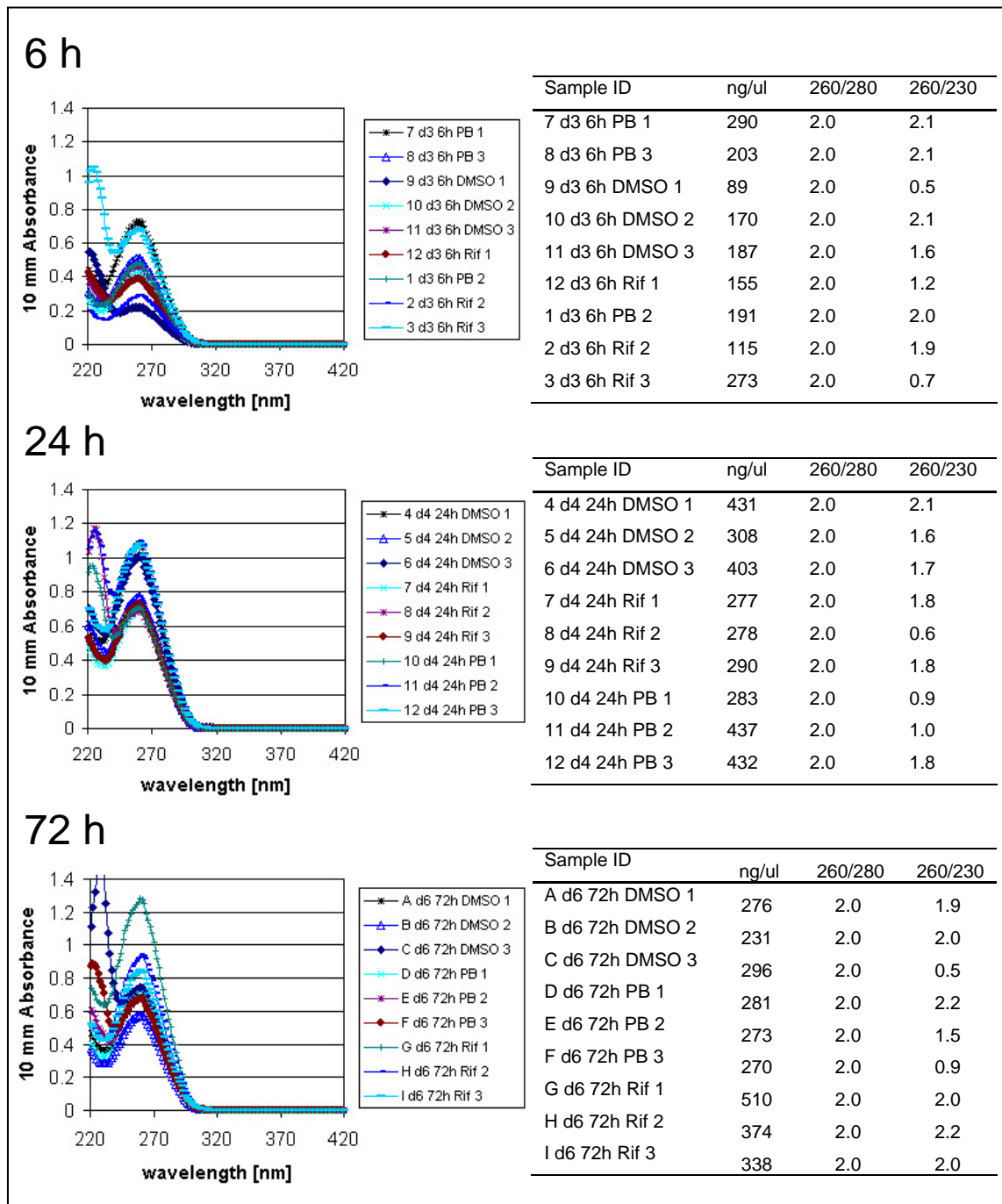


Figure 52: Absorption of RNA samples measured with the NanoDrop spectrophotometer. On the left side the absorption spectra for the RNA samples of hepatocyte cell culture HH26 6h, 24 h and 72 h. Concentrations calculated from these measurements and obtained quality estimates (260/280 and 260/230 ratio) are given on the right side of this figure.

Subsequently they were tested for their integrity by capillary gel electrophoresis. Purity is a quite important issue, however degradation state is even more important because a highly pure but degraded RNA is not suitable for further analysis. RNA integrity was ascertained by determining the RIN. The RIN stands for the quality in terms of degradation of RNA. A not degraded RNA has a RIN of 10.0 while a totally degraded RNA has < 2.0 . The threshold for good quality RNA was set to 6.5 for RNA from primary human hepatocyte cell cultures.

Figure 53 shows as an example the quality determination in concern of degradation of RNA for cell culture HH38.

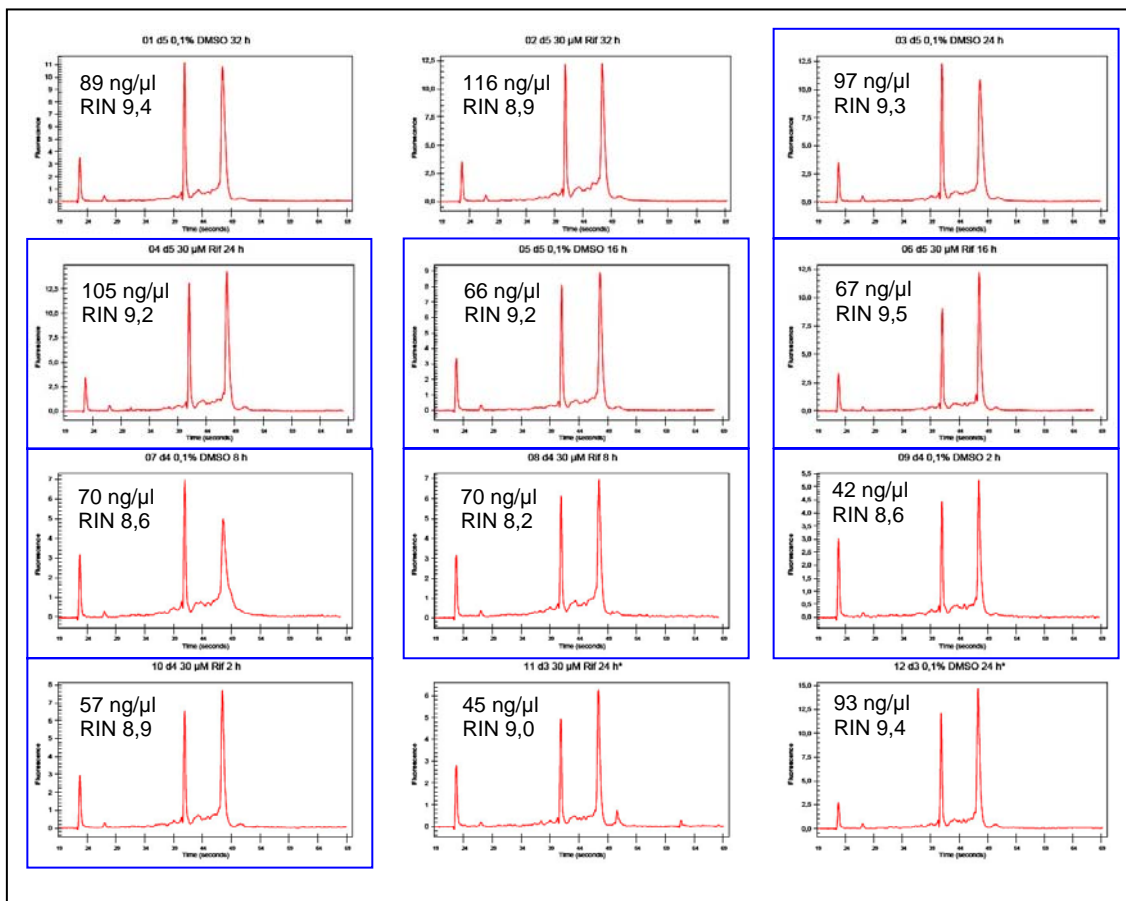


Figure 53: Electropherogram of RNA samples from human hepatocyte culture (HH38). First peak is the lower marker peak. The 5S-region (second small peak) covers the small rRNA fragments (5S and 5.8S rRNA, and tRNA). The next two peaks show the ribosomal 18S and 28S RNA. The concentration and RNA-integrity number (RIN) are shown in the diagram as well. The samples used for the microarray analysis are framed.

Plotted are the so called electropherograms which show the fluorescence of labeled RNA during the electrophoretic separation over time. Clearly visible are the not degraded ribosomal RNA (rRNA) peaks indicating not degraded mRNA as well. Also the post- and pre-ribosomal peak areas which are also basis for the RIN calculation show no indications for degradation. In all cases high quality RNA in concern of integrity was obtained. The determined RIN number ranged from 8.2 to 9.4. All samples were suitable for subsequent amplification, labeling and array analysis. The samples marked by blue boxes were hybridized on Eppendorf DualChips.

30 RNA samples from culture HH26 showed no degradation, however two samples were degraded (shown in grey in Table 30) and therefore not used for further analysis. Sample "5 d3 0h not tr. 2" had a RIN of 4.2 and for sample "2 d3 6h Rif 2" a RIN of 6.2 was determined. As both samples failed the quality criteria of $RIN > 6.5$ they were excluded from this study. Replicate samples were used to analyse this points in time and treatment.

For all samples from culture HH27 the integrity of RNA was in the required range from 8.6 to 9.6 and thus all samples from this culture were appropriate for further investigation (see Table 31).

Table 30: Amount and quality of isolated total RNA from one well of a six well plate (1.5 million cells) of culture HH26 cultivated in supplemented Williams medium E. The sample ID indicates the age of the culture (d2 = day two), the time after induction (0h, 6h, 24h, 72h) and treatment (not tr. = not treated, PB = Phenobarbital, DMSO = control, Rif = Rifampicin). The numbers given are based on measurements with the NanoDrop spectrophotometer (concentration and absorption values) and the Agilent Bioanalyzer (RIN). The required amount for an experiment (2 µg) is given as well as the number of aliquots per sample. Samples failing the necessary amount or quality are printed in italics.

Sample ID	ng/ul	260/280	260/230	RNA [µg]	RIN	volume for 2 µg [µl]	no. of aliquots
4 d3 0h not tr. 1	379	2.0	2.2	9	8.0	5.3	2
<i>5 d3 0h not tr. 2</i>	<i>457</i>	<i>1.9</i>	<i>2.3</i>	<i>11</i>	<i>4.2</i>	<i>4.4</i>	<i>4</i>
6 d3 0h not tr. 3	671	1.8	2.3	17	9.0	3.0	4
7 d3 6h PB 1	421	2.0	2.2	11	9.5	4.8	4
8 d3 6h PB 3	334	1.9	2.2	8	7.5	6.0	3
<i>9 d3 6h DMSO 1</i>	<i>55</i>	<i>1.9</i>	<i>2.0</i>	<i>1</i>	<i>6.7</i>	<i>36.5</i>	<i>1</i>
10 d3 6h DMSO 2	310	1.9	2.3	8	8.8	6.4	3
11 d3 6h DMSO 3	333	1.9	2.3	8	8.7	6.0	3
<i>12 d3 6h Rif 1</i>	<i>48</i>	<i>1.8</i>	<i>1.8</i>	<i>1</i>	<i>6.7</i>	<i>42.0</i>	<i>1</i>
<i>1 d3 6h PB 2</i>	<i>102</i>	<i>1.8</i>	<i>2.2</i>	<i>3</i>	<i>8.6</i>	<i>19.6</i>	<i>1</i>
<i>2 d3 6h Rif 2</i>	<i>120</i>	<i>1.8</i>	<i>2.3</i>	<i>3</i>	<i>6.2</i>	<i>16.7</i>	<i>1</i>
3 d3 6h Rif 3	530	1.9	2.2	13	9.0	3.8	4
4 d4 24h DMSO 1	1381	1.9	2.2	35	8.8	1.4	4
5 d4 24h DMSO 2	697	1.9	2.2	17	8.6	2.9	4
6 d4 24h DMSO 3	923	1.9	2.2	23	7.9	2.2	4
7 d4 24h Rif 1	546	1.9	2.2	14	9.0	3.7	4
8 d4 24h Rif 2	570	1.9	2.2	14	8.8	3.5	4
9 d4 24h Rif 3	548	1.9	2.2	14	7.8	3.6	4
10 d4 24h PB 1	642	2.0	2.2	16	8.1	3.1	4
11 d4 24h PB 2	968	1.9	2.3	24	8.7	2.1	4
12 d4 24h PB 3	1293	2.0	2.2	32	8.6	1.5	4
A d6 72h DMSO 1	670	1.9	2.2	17	9.3	3.0	4
B d6 72h DMSO 2	457	1.9	2.2	11	8.8	4.4	4
C d6 72h DMSO 3	746	1.9	2.2	19	9.2	2.7	4
D d6 72h PB 1	561	1.9	2.2	14	9.6	3.6	4
E d6 72h PB 2	548	1.9	2.2	14	9.4	3.6	4
F d6 72h PB 3	1014	1.9	2.2	25	8.9	2.0	4
G d6 72h Rif 1	1048	1.9	2.2	26	9.1	1.9	4
H d6 72h Rif 2	1232	2.0	2.2	31	9.2	1.6	4
I d6 72h Rif 3	766	1.9	2.3	19	8.0	2.6	4
J d6 72h not tr. 1	925	1.9	2.2	23	7.6	2.2	4
K d6 72h not tr. 2	495	1.9	2.2	12	8.0	4.0	4
L d6 72h not tr. 3	716	1.9	2.2	18	9.4	2.8	4

Table 31: Amount and quality of isolated total RNA from one well of a six well plate (1.5 million cells) of culture HH27 cultivated in supplemented Williams medium E. The sample ID indicates the time after induction (0h, 6h, 24h, 72h) and treatment (not tr. = not treated, DMSO = control, Rif = Rifampicin). The numbers given are based on measurements with the NanoDrop spectrophotometer (concentration and absorption values) and the Agilent Bioanalyzer (RIN). The required amount for an array analysis (2 μg) is given as well as the number of available aliquots per sample.

Sample ID	ng/ul	260/280	260/230	RNA [μg]	RIN	volume for 2 μg [μl]	no. of aliquots
1 0h not tr.	54	2.0	1.7	3	8.7	36.7	1
2 6h DMSO	111	2.0	2.1	5	8.6	18.0	2
3 6h RIFA	197	2.0	2.0	9	8.6	10.2	4
4 6h PB	252	2.0	2.1	12	9.0	7.9	4
5 24h DMSO	296	2.0	0.7	14	9.2	6.7	4
6 24h RIFA	301	2.0	2.2	14	9.1	6.6	4
7 24h PB	430	2.0	2.1	21	9.2	4.7	4
8 72h DMSO	251	2.0	2.1	12	9.2	8.0	4
9 72h RIFA	347	2.0	2.1	17	9.6	5.8	4
10 72h PB	258	2.0	2.0	12	9.5	7.8	4

Table 32: Amount and quality of isolated total RNA from one well of a six well plate (1.5 million cells) of culture HH44 cultivated in supplemented Williams medium E. The sample ID indicates the time after induction (0h, 6h, 24h, 72h) and treatment (DMSO = control, Rif = Rifampicin). The numbers given are based on measurements with the NanoDrop spectrophotometer (concentration and absorption values) and Agilent Bioanalyzer (RIN). The required amount for an array analysis (2 μg) is given as well as the number of available aliquots per sample.

Sample ID	ng/ul	260/280	260/230	RNA [μg]	RIN	volume for 2 μg [μl]	no. of aliquots
1 6h DMSO	311	2.0	1.5	15	n.d.	8.1	5
2 6h Rif	360	2.0	1.8	17	n.d.	6.9	6
3 24h DMSO	180	2.0	2.1	9	n.d.	13.9	3
4 24h Rif	145	1.9	1.5	7	n.d.	17.3	2
5 72h DMSO	205	1.9	1.6	10	n.d.	12.2	3
6 72h Rif	232	2.0	1.9	11	n.d.	10.8	4

3.4 Verification of Induction of Cell Cultures

Before microarray experiments were performed it was checked whether the induction of the cells was successful. As CYP3A4 is the main important enzyme in detoxification, the induction of expression of this enzyme was tested on mRNA level by real-time PCR and on protein level by western-blot.

3.4.1 Quantitative Real-Time PCR (qRT-PCR)

To prove the induction of cells an aliquot of isolated RNA was used for a CYP3A4 and a CYP2B6 TaqMan assay. The amount of CYP3A4 and CYP2B6 mRNA in Rifampicin treated and in control cells (DMSO) was determined. Furthermore the amount of 18S rRNA was measured for all samples for normalization purposes.

Based on the measurements of a reference plasmid at different concentrations a calibration curve for the CYP3A4 assay was determined (Figure 54). The coefficient of determination was 0.998. This measurement was performed to test the performance of the assay. Based on the obtained calibration curve, calculation of absolute amounts of CYP3A4 mRNA was possible. For surveying the induction, relative ratios between normalized Rifampicin treated and DMSO cells was sufficient.

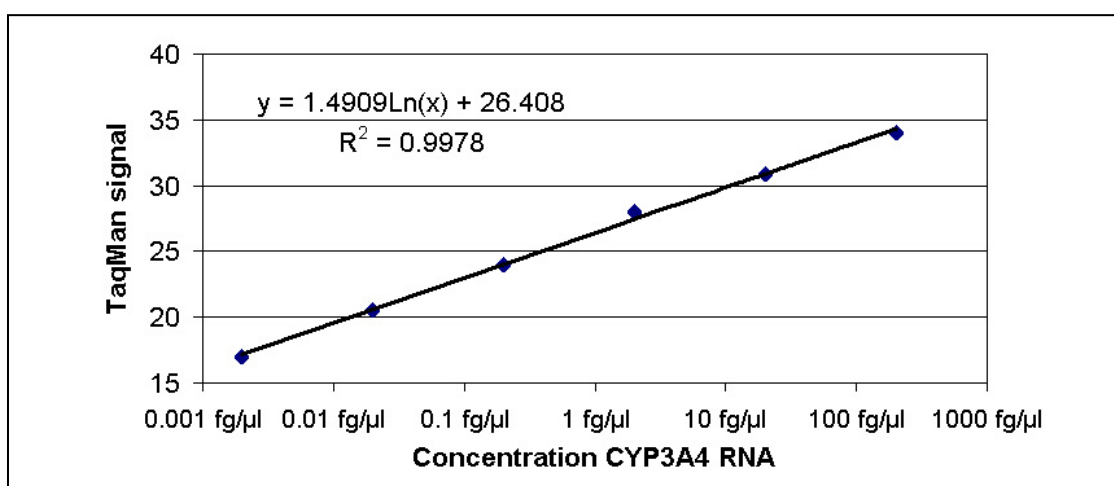


Figure 54: CYP3A calibration curve for real-time PCR assay. The plot shows the TaqMan signal measured at different CYP3A4 mRNA concentrations. Based on this calibration curve absolute amounts of mRNA in could be determined in samples.

The measurements were always performed twice as indicated in Table 33 showing the allocation of samples for a TaqMan assay performed in a 96 well plate. In this layout, one plate was used for CYP3A4, one plate for CYP2B6 and one for the S18 RNA.

Table 33: Schematic layout of a 96 well plate real-time PCR assay. Twice a calibration curve was determined (column 1 and 2) and 12 samples were measured (column 3 to 6). Dependent on the respective assay specific probes and primers were added.

	1	2	3	4	5	6
A	0.1	0.1	1	1	9	9
B	0.01	0.01	2	2	10	10
C	0.001	0.001	3	3	11	11
D	0.0001	0.0001	4	4	12	12
E	0.00001	0.00001	5	5		
F	0.000001	0.000001	6	6		
G	NTC	NTC	7	7		
H			8	8		

The average and ratios of these induction measurements were calculated using the following equations (18 - 21):

$$QTY_{\text{Rifampicin/DMSO}} = \frac{QTY_{\text{Rifampicin 1/DMSO 1}} + QTY_{\text{Rifampicin 2/DMSO 2}}}{2} \quad (18)$$

$$18S = \frac{QTY_{18S\text{-mean 1}} + QTY_{18S\text{-mean 2}}}{2} \quad (19)$$

$$CYP3A4/18S = \frac{QTY_{\text{Rifampicin/DMSO}}}{18S} \quad (20)$$

$$\text{fold over control} = \frac{(CYP3A4/18S)_{\text{Rifampicin}}}{(CYP3A4/18S)_{\text{DMSO}}} \quad (21)$$

Using equation 18, the average of the two measurements for the CYPs and 18S RNA of Rifampicin treated and DMSO control cells was determined. The varying levels of CYP3A4 and CYP2B6 RNA were normalized by the corresponding 18S RNA value, which was more or less similar in the different cell charges, as presented in Figure 55.

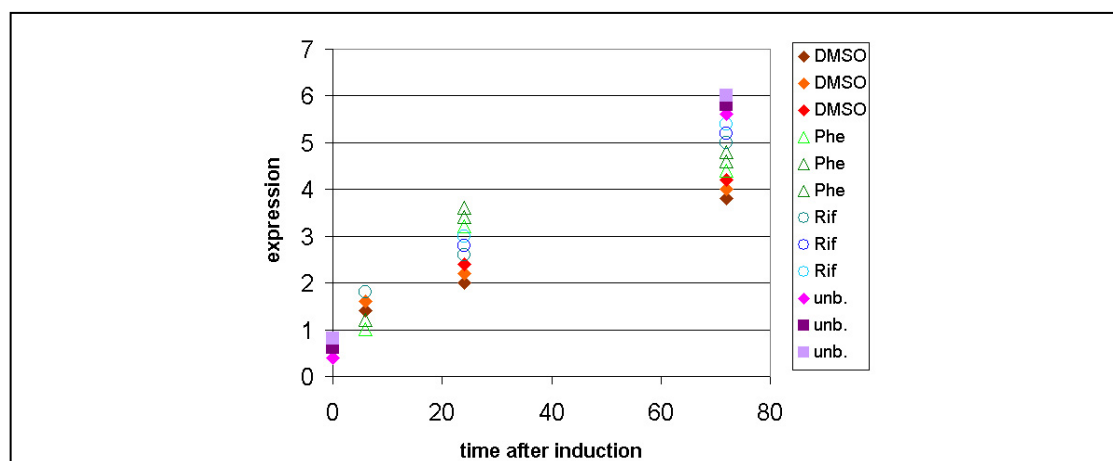


Figure 55: To normalize the expression data the content of 18S ribosomal RNA was determined which is equally and constantly expressed in the liver in different samples. The variability of expression is increasing with the duration of the cultivation. The replicate measurements fit together and even at 72 h where the highest variation occurred, normalization to 18S expression level was possible.

Finally the ratio between normalized RNA level in Rifampicin treated cells and normalized RNA level in DMSO control cells was calculated (equation 21) resulting in a fold induction over control estimate. The results of these measurements are summarized in Figure 56.

The CYP3A4 mRNA levels shown in A and B were significantly up-regulated. The induction with phenobarbital (as positive control) was about 3 fold after 24 h and 9-fold after 72 h. One biological replicate of the three measurements showed a smaller induction (4-fold) after 72 h. The Rifampicin induction of CYP3A4 was a bit lower (part B of Figure 56). After 24 h a 3.5-fold induction could be observed. After 72 h of induction the effect on the three different biological replicates was different, ranging from about 2- to 4-fold induction (part B of Figure 56).

The expression level of CYP2B6 was also increased in phenobarbital and Rifampicin treated cells compared to similar cells treated with DMSO (control). The induction of CYP2B6 with phenobarbital (part C of Figure 56) was about 5-fold after 24 hours and 11- to 12-fold after 72 hours. After the 72 h again a higher variance was observed

between replicates. CYP2B6 was also induced by Rifampicin, showing a 7- to 9-fold induction after 24 h and a 5- to 12-fold induction after 72 h. The induction of CYP2B6 at 72 h showed the highest variability between the biological replicates. This phenomenon could be observed in all samples. The variability of expression levels increased over time. After 6 h only slight, up to 2-fold induction with phenobarbital was observed.

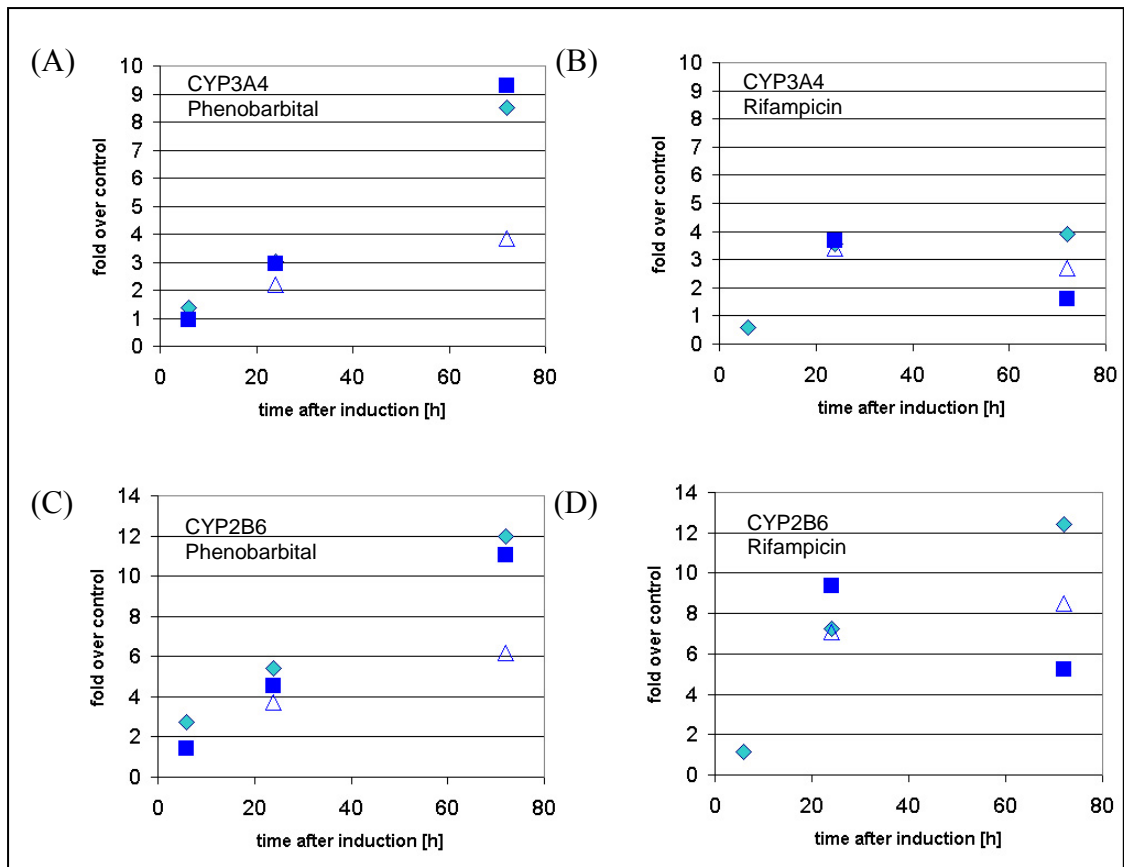


Figure 56: Verification of induction of primary human hepatocyte cell cultures using real-time PCR measurement. The induction of two Cytochrome P450s (CYPs) was assessed to proof successful induction. Shown here are two examples of the standardized (against control cells) fold change of CYP3A4 (A and B) normalized to 18S RNA and induction of CYP2B6 (C and D) normalized on 18S RNA. In the first row (A and C) cells were induced with phenobarbital resulting in a 9- to 10-fold induction and in the second row using Rifampicin (B and D) a 4- to 12- fold induction was monitored. Samples were taken and analysed after 6, 24 and 72 h.

3.4.2 Western Blotting

To confirm the induction of CYP3A4 by Rifampicin not only on mRNA level using real-time PCR, flow through of the RNA isolation was precipitated using acetone as mentioned in 2.9.1. The protein containing precipitate was separated by SDS-PAGE (2.9.2) and finally analysed by western blot (2.9.3). An example of the results is shown in Figure 57. In concordance with the real-time PCR measurements no induction was observed on protein level after 6 h. After 24 h a band of the molecular weight of CYP3A4 appeared in both samples (Rifampicin and DMSO). The band in lane five (Rifampicin) is stronger than the band of the same size in lane four (DMSO). The difference between DMSO (control) and Rifampicin is much stronger after 72 h indicating a significant induction due to Rifampicin administration. In agreement with the real-time PCR results an induction of the primary cell cultures could also be observed on protein level. Although the strongest induction for CYP3A4 was determined after 72 h at protein level while on mRNA level the strongest induction was detected after 24 h.

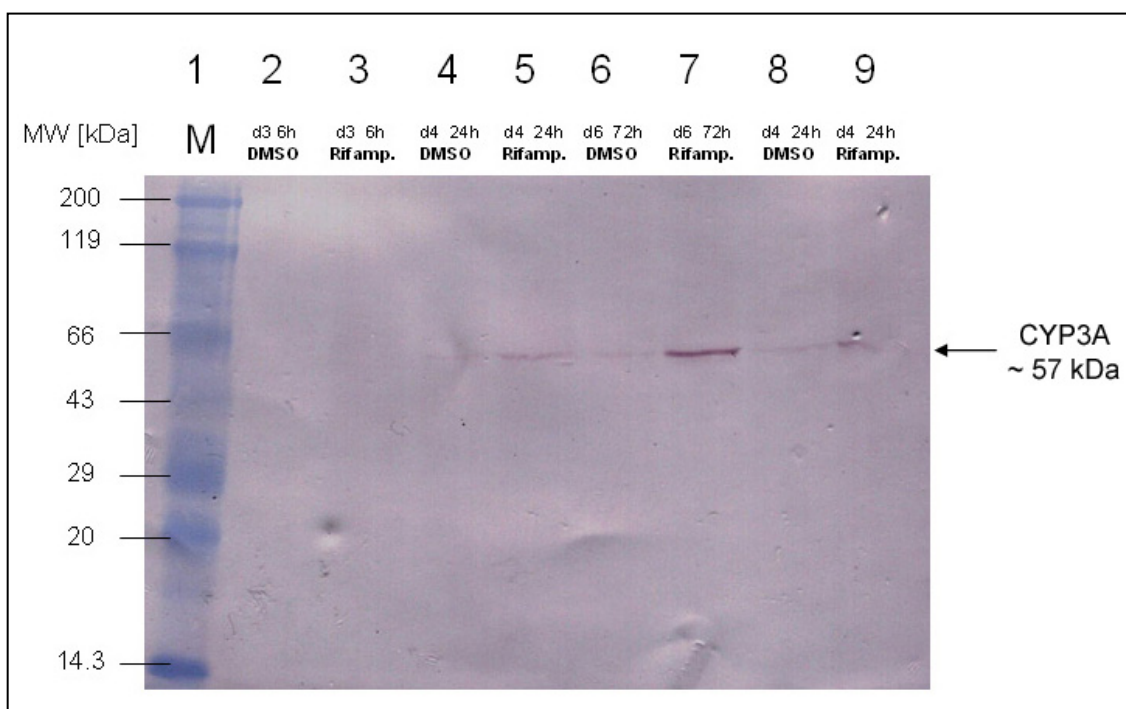


Figure 57: CYP3A Western-Blot from protein containing flow through of RNA-isolation procedure. The gel was blotted and the blot incubated with CYP3A antibody (WB-MAB-3A Human CYP3A Western Blotting Kit, No. 458254 lot 21212 from Gentest, BD Biosciences).

In all tested samples (HH26, HH27, HH44) successful induction with Rifampicin was measured. All samples were used for analysis of the expression levels of a larger set of genes based on microarray studies.

3.5 Reverse Transcription (RT), Amplification and Labeling of mRNA

After determining the quality of the isolated RNA and verifying the induction using qRT-PCR and Western Blot the samples had to be prepared for microarray analysis. There is a variety of different labeling methods and protocols available which require 0.5 to 25 μg of total RNA template. Usually this amount can be obtained from roughly 100,000 to 5,000,000 mammalian cells. 1.5 mio human liver cells yielded approximately 5 to 35 μg RNA.

There are different approaches for labeling on the market like chemical- or enzymatic methods or direct- and indirect labeling. Chemical methods like *cis*-platinum derivatives seem to be quite promising however are not yet widely used. In our first approach we stuck to the classical enzymatic labeling approach which involves the direct incorporation of Cy dye modified nucleotides into cDNA (Schena *et al.*, 1995).

3.5.1 Direct Labeling with Fluorescent Dyes

Using the classical enzymatic approach cDNA is produced by reverse transcriptase in a reaction with an oligo dT-primer (2.8.9). Based on the most recent protocols 20 - 100 μg of total RNA should be used for labeling. During the reverse transcription reaction modified nucleotides are incorporated into the cDNA. The modification is often fluorescence labeling, allowing the direct detection of the cDNA on the microarray by scanning with a fluorescence scanner. For the reverse transcription reaction 10 μg of total RNA were used yielding in 20 to 30 μg of labeled cDNA. The incorporation of cy-dye nucleotides could be verified by measuring the absorption at 550 nm for Cy3 and 649 nm for Cy5. A spectrum of labeled cDNA sample is given in Figure 58.

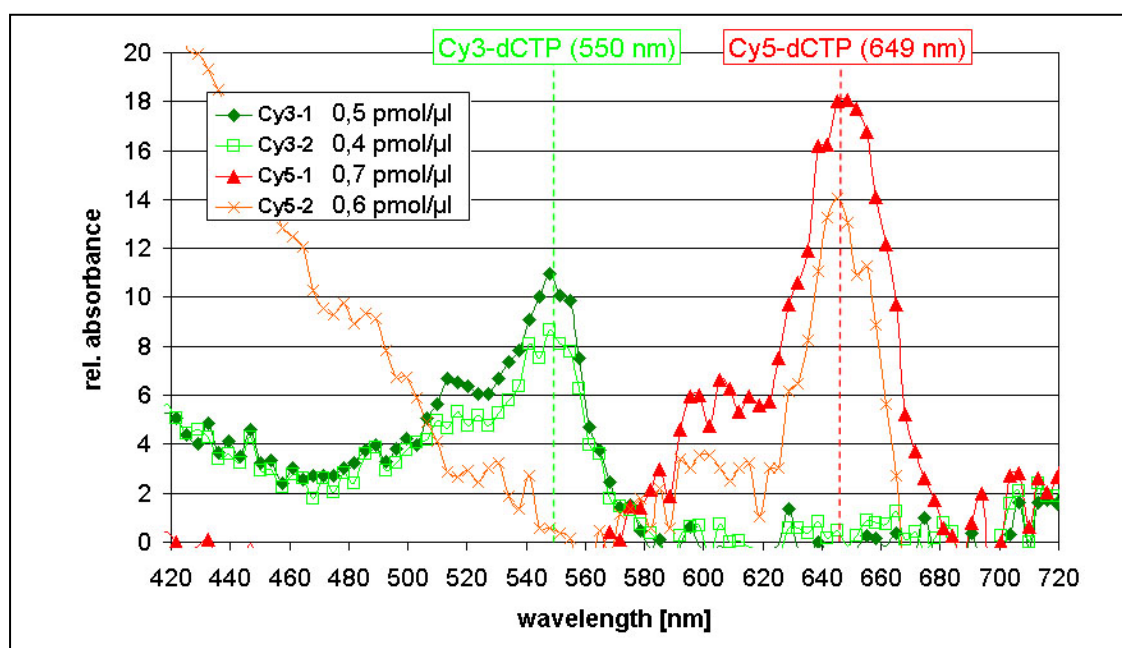


Figure 58: Absorption of labeled cDNA for hybridization on microarrays measured with NanoDrop spectrophotometer. The fluorescence dyes (Cy3 and Cy5) were incorporated into the cDNA during reverse transcription. Based on these measurements the incorporation rate of dye/nucleotide can be calculated.

During reverse transcription 30 - 60 ng/ μ l labeled cDNA were generated which were then hybridized on the Eppendorf DualChips. An example of the results for reverse transcription and labeling is given in Table 34.

Table 34: Amount of incorporated fluorescent dye during reverse transcription reaction. The amount of Cy-dye, the concentration of cDNA and the ratio of 260/280 nm were determined using the NanoDrop spectrophotometer.

Sample ID	Cy3 [pmol/ μ l]	Cy5 [pmol/ μ l]	cDNA [ng/ μ l]	260/280
Cy3-1	0.5	0.0	34.7	1.5
Cy3-2	0.4	0.0	31.6	1.6
Cy5-1	0.0	0.7	62.2	1.9
Cy5-2	0.0	0.6	58.5	1.6

Using the amount of fluorescence measured and the concentration of cDNA the incorporation rate of the sample could be calculated. The molecular mass of one nucleotide is 330 g/mol. The incorporation rate is given by the following equation (22):

$$\text{incorporation rate} = \frac{\text{DNA concentration } \left[\frac{\text{g}}{\mu\text{l}} \right]}{\text{molecular Mass / nucleotide } \left[\frac{\text{g}}{\text{mol}} \right]} \cdot \frac{1}{\text{amount of Cy } \left[\frac{\text{pmol}}{\mu\text{l}} \right]} \quad (22)$$

Measured concentrations for incorporated fluorescence dyes (Cy3 and Cy5) in pmol/ μl and DNA-concentration in ng/ μl using NanoDrop spectrophotometer were:

Cy3	0.5 pmol/ μl	Cy5	0.7 pmol/ μl
DNA	33 ng/ μl	DNA	60 ng/ μl

DNA concentration was divided by the molecular mass of one nucleotide to obtain the molarity.

$$\frac{\text{DNA concentration } \left[\frac{\text{g}}{\mu\text{l}} \right]}{\text{molecular Mass of one nucleotide } \left[\frac{\text{g}}{\text{mol}} \right]} = \frac{33 \cdot 10^{-9} \frac{\text{g}}{\mu\text{l}}}{330 \frac{\text{g}}{\text{mol}}} = 1^{-10} \frac{\text{mol}}{\mu\text{l}} \quad \left[\frac{\text{g}}{\mu\text{l}} \cdot \frac{\text{mol}}{\text{g}} = \frac{\text{mol}}{\mu\text{l}} \right]$$

Based on these measured concentrations the incorporation rate of dyes per nucleotide could be calculated using equation 22:

$$\frac{1^{-10} \frac{\text{mol}}{\mu\text{l}}}{0,5 \cdot 10^{-12} \frac{\text{mol}}{\mu\text{l}}} = \underline{\underline{200}} \quad \text{1 Cy3-dye molecule per 200 nucleotides}$$

$$\frac{1,82 \cdot 10^{-10} \frac{\text{mol}}{\mu\text{l}}}{0,7 \cdot 10^{-12} \frac{\text{mol}}{\mu\text{l}}} = \underline{\underline{260}} \quad \text{1 Cy5-dye molecule per 260 nucleotides}$$

For sample Cy3 with an average concentration of 33 ng/ μl cDNA and 0.5 pmol/ μl Cy-dye (measured twice, given in Table 34), an incorporation rate of 1 Cy3 per 200 nucleotides was calculated. The sample Cy5 consisting of 60 ng/ μl contained 0.7 pmol/ μl Cy-dye (measured twice, given in Table 34) resulting in one Cy-dye per 260 nucleotides.

The amount of RNA from primary cell cultures was very limited and the necessary amount for this kind of labeling is quite high (20 - 100 μg). For this reason and due to concerns about "dye bias" another strategy had to be taken into account. The reason for this so called "dye bias" is dye-specific differences in the incorporation of modified nucleotides into cDNA. There are different strategies to avoid this kind of bias. On one hand aminoallyl-modified bases can be incorporated into DNA, followed by Cy3 and Cy5 coupling to the reactive amino groups. Another approach of this "indirect" labeling method is the use of biotinylated nucleotides. This version of the technique in combination with amplification, also adopted by Affymetrix (Dalma-Weiszhausz *et al.*, 2006), was used to overcome the problem of dye bias, as well as low amount of total RNA from primary cell culture.

3.5.2 Amplification and Indirect Labeling Using Biotin Modified Nucleotides

To obtain higher amounts of nucleic acids, especially for replication and verification of experiments, amplification techniques had to be applied. One prerequisite for amplification with subsequent expression profiling is that it works without altering the complexity and composition of the original RNA pool. The method chosen for this purpose was so called *in vitro* transcription (IVT), based on a linear amplification with T7 RNA polymerase. Nowadays several suppliers offer complete kits for this purpose. In this work the MessageAmpTM II aRNA Kit from Ambion was used for amplification and labeling of total RNA (2.8.5).

The amplification was monitored by measuring the absorption in a spectrophotometer to calculate the obtained concentration as well as by determining the size distribution of amplified fragments using capillary gel electrophoresis. An electropherogram and an *in-silico* gel like image of an amplified sample are shown in Figure 59.

In this electropherogram (A) the fluorescence of the labeled RNA is given over time as it is separated during capillary electrophoresis. Shown as red curve, the size of amplified fragments was determined to be from 0.1 kB to 7.0 kB. A size standard (RNA ladder from Ambion RNA6000) with the fragment sizes 0.2, 0.5, 1.0, 2.0, 4.0, 6.0 kB is shown in light blue, plotted against the secondary y-axis. The starting RNA material,

with its two characteristic ribosomal RNA peaks indicating no degradation and high quality is shown as orange line.

Part (B) of Figure 59 shows a gel like image obtained from the same sample after separation using capillary gel electrophoresis. Lane one shows the size marker (ladder) from Ambion (RNA 6000). In the second lane the source RNA sample (RNA 9) is given, showing the two ribosomal RNA peak bands. The third lane shows the amplified RNA sample with amplification products from about 0.1 kB to 7.0 kB. The strongest amplification was determined for fragments between 0.5 to 3.0 kB. In the example presented in Figure 59 1.4 μg of RNA (RNA 9, 477 $\text{ng}/\mu\text{l}$) were used as starting material. A final concentration of 1320 $\text{ng}/\mu\text{l}$ was obtained, corresponding to 130 μg . The amplification factor was 90-fold. Amplification factors between 40 and 180 were achieved for different samples.

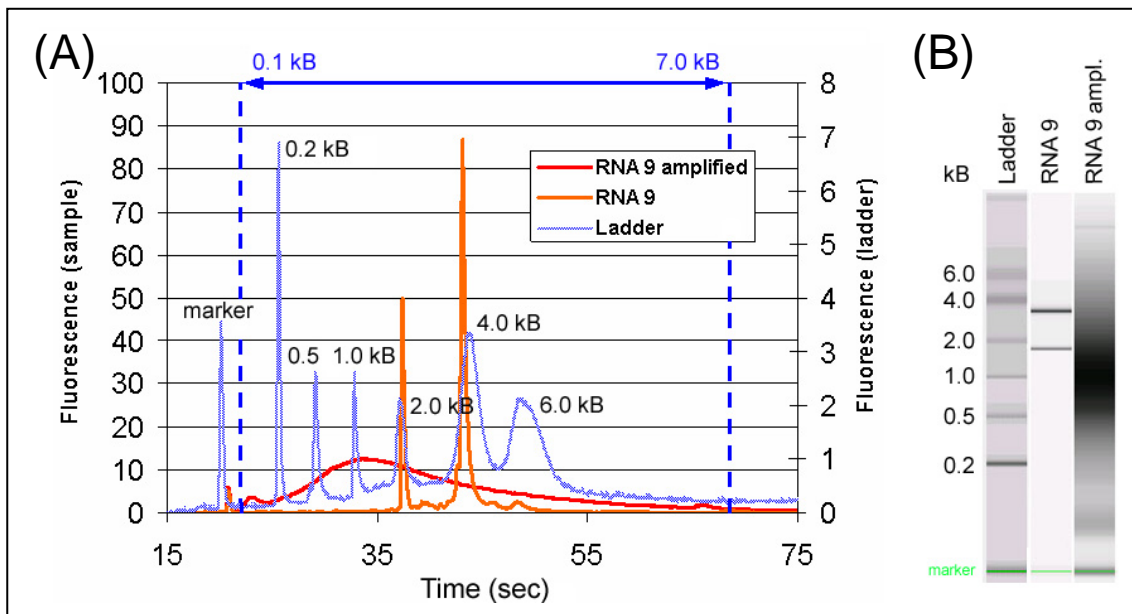


Figure 59: Agilent Bioanalyzer run showing the results of analysing amplified (aRNA) and not amplified mRNA samples. In part (A) the electropherogram of a not amplified sample is given in orange and shows the expected two ribosomal peaks of intact degraded RNA sample. The amplified RNA is given in red and shows a fragment size distribution from 0.1 kB to 7.0 kB of amplified products. Part (B) shows a gel-like representation of the measured absorbance. Lane one of this image present a RNA size marker (RNA6000 from Ambion), the second lane shows the total RNA used for the amplification and lane three gives the results of this amplification.

Amplification of test and reference samples was always reproducible and resulted in all cases in amplification of fragments with the size from 0.2 to about 6.0 kB. Ten amplified samples from HH26 are shown in Figure 60 as gel-like image and electropherograms in Figure 61. A similar size distribution was found in all samples. To allow comparison of not amplified RNA with amplified product of the reaction, not amplified total RNA was added to the assay in lane 9 and 11.

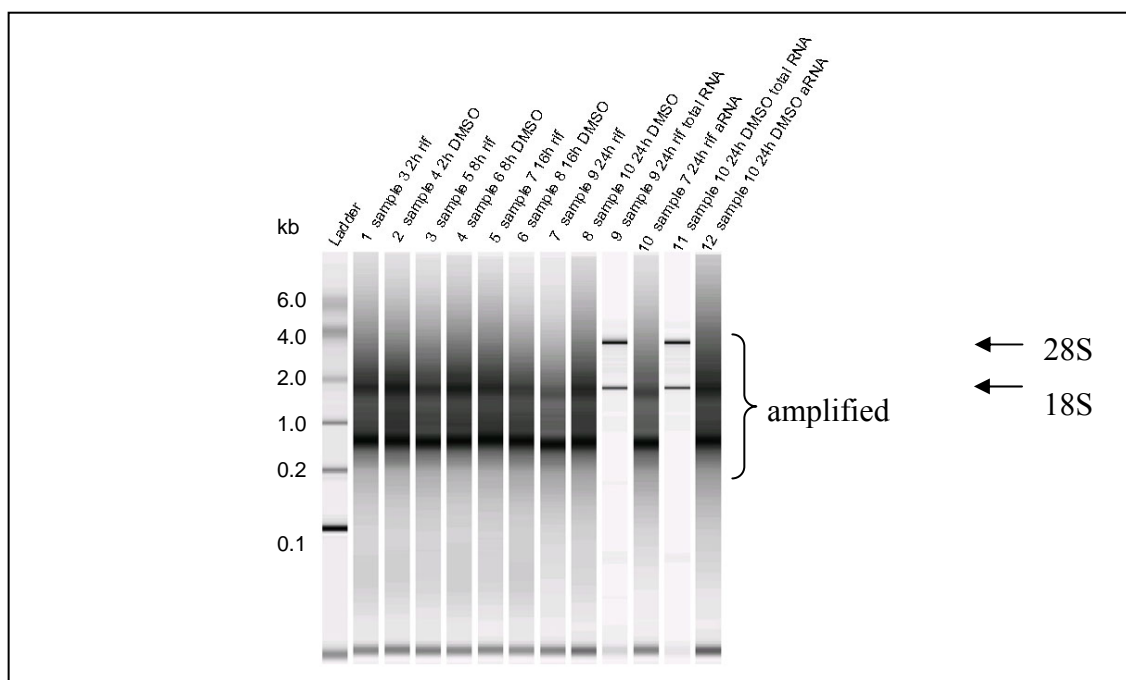


Figure 60: Gel like image of the amplified aRNA samples for microarray analysis. Lane 1 to 8 show amplified RNA samples. The aRNA was obtained after *in vitro* transcription from total RNA of induced and not induced primary human hepatocytes. In lane 9 and lane 11 the starting material (total RNA) is shown. The two ribosomal bands, 1.9 kb for 18S-RNA and 3.7 kb for the 28S-RNA are clearly detectable. In lane 10 and 12 the resulting aRNA after amplification from samples 9 and 10 (lane 9 and 11) can be seen.

0.5 μg of RNA were used as starting material for amplification of samples HH26. The concentration of total RNA was between 62 and 117 $\text{ng}/\mu\text{l}$. Therefore 4.3 to 8.1 μl of RNA had to be used. The amplification yielded 30 to 50 μg of RNA, corresponding to an amplification factor of 60 to 100. A priori the labeling rate with biotinylated nucleotides could not be determined. The aRNA was measured by NanoDrop and Bioanalyzer to determine the concentration and prove amplification of short as well as long RNA fragments. Successfully amplified samples (as shown in Figure 61) were then hybridized on the DualChip microarrays.

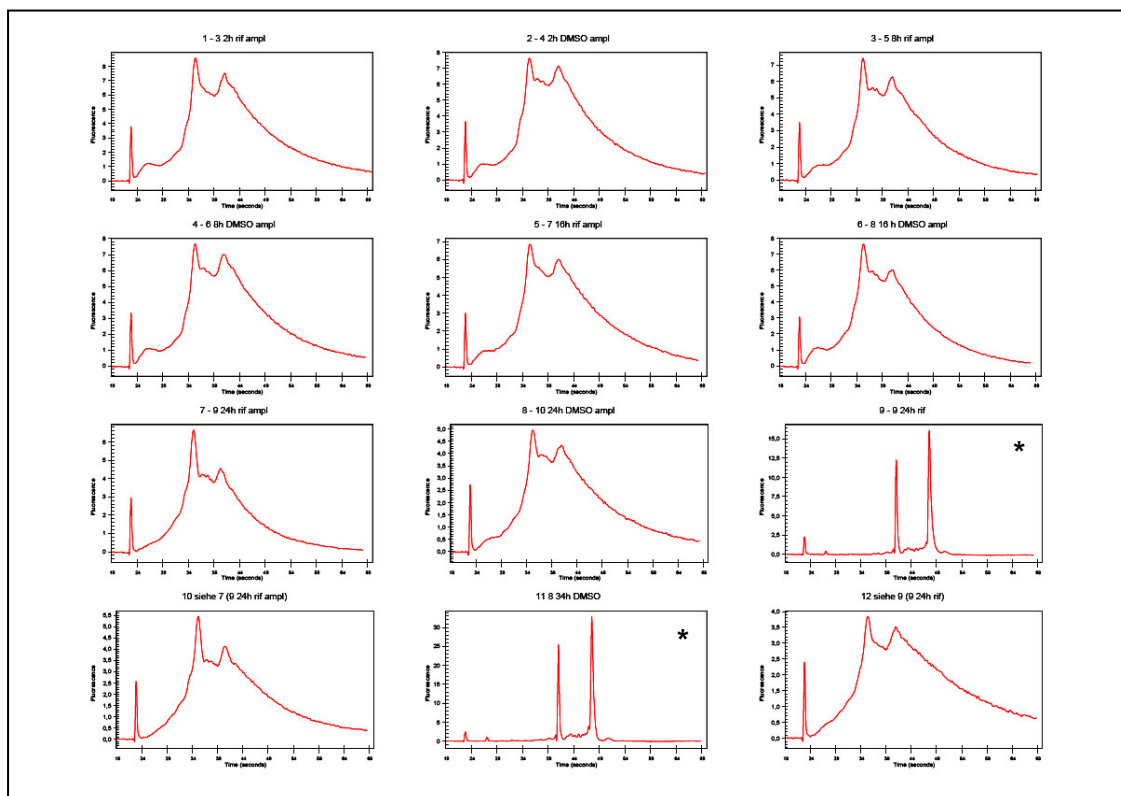


Figure 61: Electropherogram of amplified RNA (aRNA) samples from human hepatocyte RNA. The amplified aRNA samples show an amplification of all RNA species of different length. As controls two not amplified mRNA samples were measured (*). These samples show the expected spectra for the two not degraded ribosomal RNA bands.

Table 35: This table gives the amount, concentration, quality and amplification factor of amplified aRNA samples. The first column indicates the denotation of samples, the second gives the name, point in time and treatment. The third column (ng/ μ l) shows the concentration of the total RNA (input). The fourth column shows the amount of sample obtained from 0.5 μ g starting material for amplification. The following columns show the concentration, purity, total amount and amplification factor of aRNA samples. The last column shows the volume in μ l, corresponding to 5 μ g of RNA used for the hybridization on the DualChips™ from Eppendorf.

no.	sample	ng/ μ l	[μ l] for 0,5 μ g	ng/ μ l	260/280	260/230	after amp. [μ g]	amplification	for 5 μ g [μ l]
3	2 h rif	79	6.3	493	2.12	1.81	46.9	94	5.1
4	2 h DMSO	62	8.1	462	2.12	1.68	43.9	88	5.4
5	8 h rif	108	4.6	377	2.13	1.7	35.8	72	6.6
6	8 h DMSO	83	6.0	521	2.11	1.86	49.5	99	4.8
7	16 h rif	97	5.1	453	2.13	1.21	43.1	86	5.5
8	16 h DMSO	104	4.8	436	2.14	1.91	41.4	83	5.7
9	24 h rif	113	4.4	324	2.13	1.66	30.7	61	7.7
10	24 h DMSO	117	4.3	335	2.14	1.04	31.8	64	7.5

3.6 Hybridization of Arrays and Quality Control

The amplified and labeled samples were used for hybridization on Eppendorf DualChips. The amplification and labeling for the full genome chips from Affymetrix was contracted to the Affymetrix facility in Tübingen. The applied method there uses a similar amplification principle to the one used for the other arrays in this study (Van Gelder *et al.*, 1990; Dalma-Weiszhausz *et al.*, 2006). The hybridization, washing, scanning and quantification were performed as described in (2.10.5, 2.10.6, 2.10.7 and 2.10.8). The success of this process was verified before proceeding to a detailed analysis of the microarray data. In the following part, a summary of the quality check for array performance is given. This section is divided into two parts. The first part (3.6.1) is dealing with quality tests for the sub-genome arrays from the company Eppendorf and the second part (3.6.2) gives the results of quality determination for the full genome chips from Affymetrix.

3.6.1 Quality Control for Sub-Genome Arrays from Eppendorf

Quality of the sub-genome arrays from Eppendorf was determined by several controls implemented on the arrays. In addition the total number of genes identified on the array was used as a marker for quality of hybridization and hybridized sample. Table 36 gives an example of such an analysis. In this example four not induced (control) arrays (10023, 10044, 10041 and 10039) of a time series experiment were investigated. In total between 317 and 342 genes were detected on these arrays. This corresponds to 77 - 83 % of all genes on the chip, as shown in Table 37. Coverage of more than 65% was estimated as successful hybridization of good quality RNA. Furthermore the control spots had to be taken into consideration. The detection curve concentration (DCC) gives a clear answer on how successful the detection was performed using indirect labeling. The signal strength of this control is only dependent from the detection. It is independent from the sample quality and labeling. On the detection curve concentration biotin is spotted in different concentrations on the array. During detection with the cy3-

conjugated anti-biotin antibody this concentration series is labeled by the dye-modified antibody.

Table 36: Number of spots detected on the not induced arrays (control) for four points in time. If a probe has a sufficiently strong result in two PMT settings it was called present. If it occurred only at one PMT setting it was called absent and is not shown here.

Arrays not induced	10023	10044	10041	10039
Samples	2h	8h	16h	24h
DCC	28	27	28	27
Gene	317	318	342	327
HKG	39	38	39	39
NDC	0	0	0	4
NHC	0	1	0	2
PHC	12	12	12	12
Standard	36	36	36	36

Table 37: Percentage of detected spots on the not induced arrays (control). If a probe has a sufficiently strong result in two PMT settings it was called present. If it occurred only at one PMT setting it was called absent and is not shown here. Around 80 % of the genes were detected. Housekeeping genes were almost completely detected. Almost no negative control was detected, only a few on array 10039. The positive hybridization control worked quite well and also the standards were completely detected. A positive signal for 90 % of the detection curve could be observed.

Arrays not induced	10023	10044	10041	10039
Samples	2h	8h	16h	24h
DCC	93 %	90 %	93 %	90 %
Gene	77 %	77 %	83 %	79 %
HKG	100 %	97 %	100 %	100 %
NDC	0 %	0 %	0 %	39 %
NHC	0 %	7 %	0 %	13 %
PHC	100 %	100 %	100 %	100 %
Standard	100 %	100 %	100 %	100 %

27 to 28 spots of the detection curve were detectable which corresponds to 90 %. This result was absolutely satisfying. The housekeeping genes which are used also for normalization were detected on all arrays for all samples with almost 100 % (39 genes). Only on one array (10044 8 h) one housekeeping gene was absent. The negative detection control (NDC) gave four positive results (39 %) for one array (10039 24 h). This is a high value compared with the other arrays. All other arrays showed no signals for the NDC. The second negative control (negative hybridization control - NHC) was present on two arrays (10044 and 10039). Array 10044 had one positive signal for the

NHC (7 %) and array 10039 two positive signals for the NHC (13 %), indicating again some background problems for array 10039. The positive hybridization controls (PHC, 12) and the internal standard (36) were detected to 100 % on all arrays for all samples. As the two positive signals for the negative controls on array 10039 were below 40 % this array was still included in the subsequent data analysis.

Similar analysis was performed for the test arrays. The results match the quality thresholds just as for the not induced arrays and are summarized in Table 38. The DCC was detected to 90 %, the total amount of genes ranged between 66 and 73 %. Housekeeping genes were detected by about 95 %. The negative controls showed some false positive signals. For the negative detection control 1% of the signals were detectable and for the negative hybridization control in two cases no signal (as expected) was observed (10023 and 10039), in one case (10044) 7 % were detected and in one case (10041) 20 % were found. The positive hybridisation control and the internal standard were identified correctly by 100 %. All arrays presented here could be used for further analysis. Arrays failing the quality test had to be repeated till all quality parameters were fulfilled.

Table 38: The number of spots detected on the induced arrays (test) for four points in time is given in this table. If a probe has a sufficient strong result in two PMT settings it was called present. If it occurred only at one PMT setting it was called as absent and is not shown here.

Arrays induced	10023	10044	10041	10039
Samples	2h	8h	16h	24h
DCC	90 %	90 %	90 %	90 %
Gene	69 %	66 %	73 %	69 %
HKG	92 %	95 %	97 %	95 %
NDC	1 %	1 %	0 %	1 %
NHC	0 %	7 %	20 %	0 %
PHC	100 %	100 %	100 %	100 %
Standard	100 %	100 %	100 %	100 %

Indirect labeling was mainly used in this study as it is more sensitive and has not to deal with dye bias. As it is dependent on a reliable detection step after hybridization, several antibody charges were tested for their storage stability and overall performance in detecting biotin. In Figure 62 the results of this study are summarized.

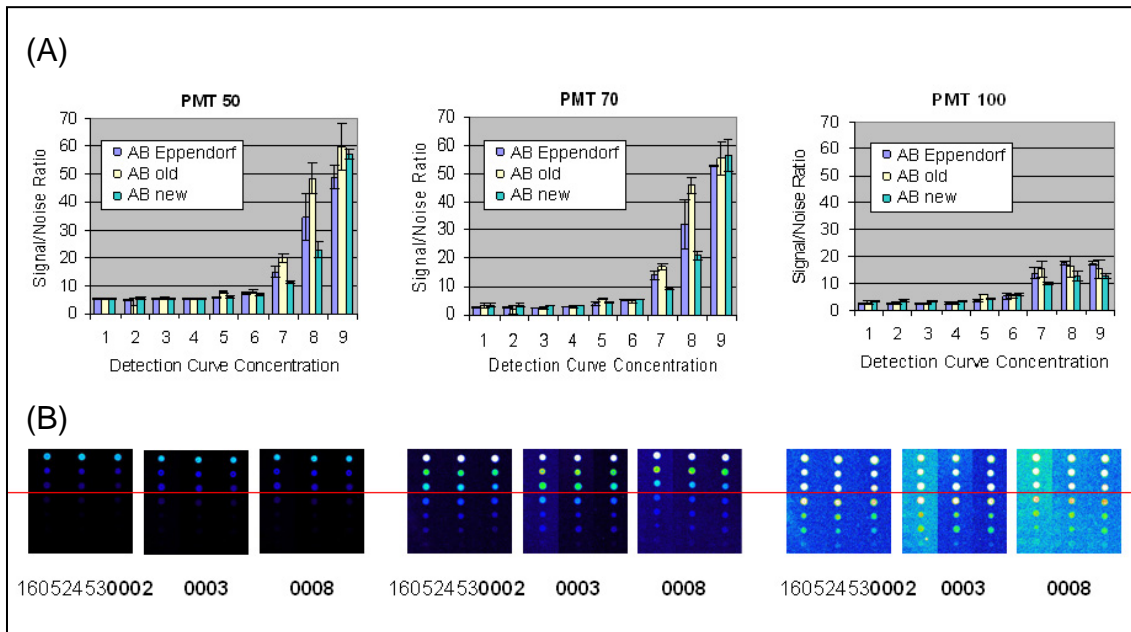


Figure 62: This figure shows the performance of different anti-biotin antibody charges, detecting different concentrated biotin spots on three microarrays (0002, 0003, 0008). The biotin detection curve concentration is diluted from 1. 1:3000; 2. 1:600; 3. 1:300; 4. 1:60; 5. 1:30; 6. 1:12; (red line); 7. 1:6; 8. 1:3; 9. not diluted. (A) The three diagrams show the signal to noise ratio at different photomultiplier settings (PMT) for the microarray laser scanner. The different antibody charges were comparable while the new diluted one showed a bit lower performance than the 1 year old one and the one used at Eppendorf. For PMT 50 and 70 the best results were obtained while at PMT 100 the background fluorescence was strongly increasing leading to a reduced signal to noise ratio. (B) The strong increase of the background fluorescence at high PMT settings can also be seen in the pictures from the scanned detection curve concentrations on the microarrays.

Three different anti-biotin antibody charges were tested. The one used at Eppendorf (as positive control), one that was used already for one year and one that was new and just prepared for this test. The detection was successful for all three antibody charges, at three different photomultiplier settings of the microarray scanner. For the highest concentration of biotin all antibodies showed more or less the same results. At lower concentrations the newly prepared one had a less intensive signal. Nevertheless it was still clearly visible over the background. The detection limit at low PMT settings (50) was at the 1:12 dilution indicated by a red line in the section of the array (Figure 62 B). The detection of a higher dilution at higher PMT settings was possible (as shown in Figure 62 B). Due to higher background the signal to noise ratio declined (Figure 62 A).

3.6.2 Quality Control for Full-Genome Arrays from Affymetrix

The chip-wide statistics tool implemented in the "refiner" software from Genedata allows computing of average intensity, percentage of outliers, percentage of masked features and the chip-wide scaling factor for the condensed values. These activities were used to quickly scan experiments for identification of arrays and features having unexpected values in any of the statistics. By assessing the values for the masked area (%), outlier area (%), and the average intensity of the features on the chip the identification of experiments with unacceptable hybridization quality was done.

Affymetrix Statistical (MAS5/GC-RMA) activity presents a summary table of the overall quality classification for each chip (Figure 63). The traffic light system provides an immediate overview of data quality corresponding to each chip. Green colours stand for good results yellow for medium and red for poor quality criteria. The overall classification reflects a total quality assignment which condenses the individual quality measures. 90 % (27) of our arrays showed a good overall quality and 10 % (3) were of medium quality.

The masked area and the outlier area are thresholds that determine the quality rating of a chip, based on the percentage of masked or outlier features. The summary for the 30 full genome chips of this study are presented in Figure 63. The percentage of masked and outlier area was for all chips below 0.2 %. Therefore all chips were classified "good" (green).

The 3'/5' ratio is another quality check, which allows determination of the level of degradation. High 3'/5' ratio indicates a difference in amplification between the 5' end and the 3' end which might arise as a result of degradation. Samples with a 3'/5' ratio of below 3.5 were considered suitable.

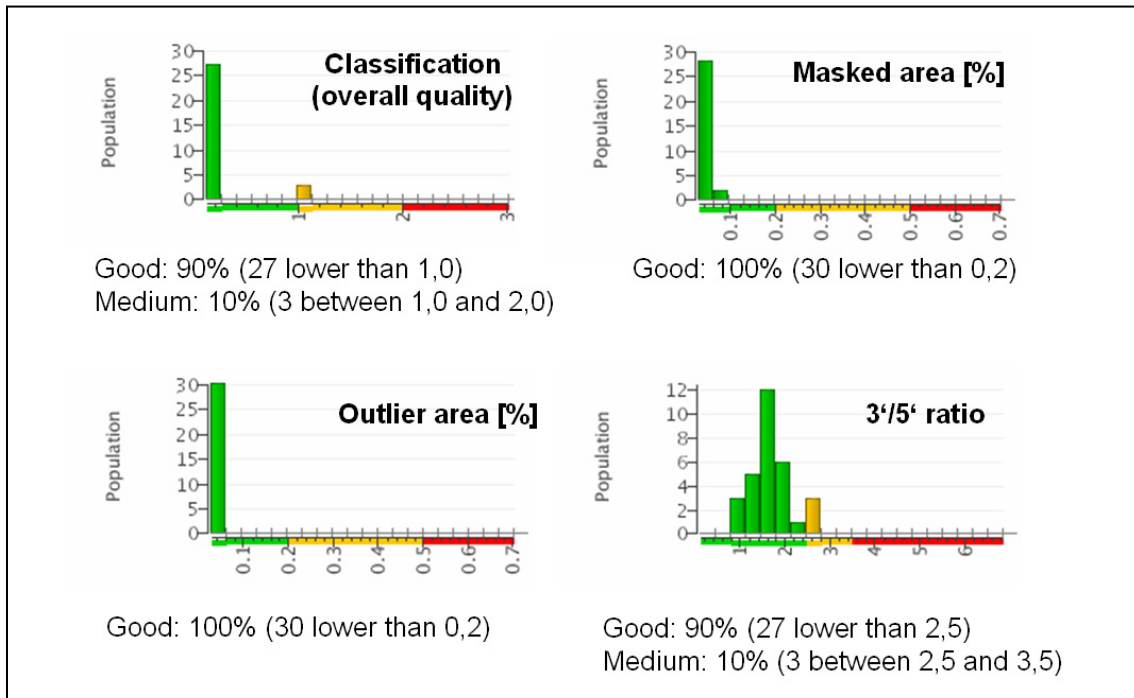


Figure 63: Quality control of Affymetrix Chips based on masked area, outlier area and 3'/5' ratio. The overall quality is given in the first graph "classification". 90 % of all arrays were classified as "good" and 10 % medium, due to slight degradation of the RNA sample indicated by a reduced (medium) 3'/5' ratio.

The average intensity computed as the arithmetic mean of all feature intensities including controls and empties is also a hint for good hybridization quality. It was checked if there were single experiments with an average intensity far off from the other experiments. The median background was 57 which is in the normal range, median scaling factor was between 1.0 and 2.3, and mean house-keeping gene 3'/5' ratio was 1.7 with a 3'/5' standard deviation of 1.1, indicating a high overall array and RNA quality.

3.7 Investigation of Technical- and Biological Replicates

To identify the variance of the full genome microarray platform, three technical replicates of cell culture HH26 were performed. The cells were cultivated and induced separately. The preparation, amplification and labeling as well as the hybridization were also done individually for each replicate. Finally the obtained fluorescence intensity of the three different arrays for each treatment and sampling point were compared. The expression profiles for five genes are given in Figure 64. It can be seen that the first three technical replicates from cell culture HH26 (the first three spots representing the array signal plotted on the x-axis blue = DMSO and red = Rifampicin) show a similar expression pattern. Cytochrome P450 3A4 (CYP3A4) had a mean expression value of 1826 ± 150 for DMSO and 7563 ± 300 for Rifampicin treated samples. For the 5-aminolevulinic acid synthase (ALAS1) a similar result was obtained with 1059 ± 15 for DMSO and 2796 ± 29 for Rifampicin. For lower expressed genes, like the transporter ATP-binding cassette sub-family B member 1 (ABCB1) and cytochrome P450 2C19 (CYP2C19), the variance was about the same as for the high expressed genes. ABCB1 showed an average expression level of 468 ± 13 (DMSO) and 1055 ± 33 (Rif) and CYP2C19 had 405 ± 52 (DMSO) and 992 ± 181 . The technical variance had a high impact on low expressed genes like nuclear receptor 1I3 constitutive androstane receptor (NR1I3, CAR). The expression value of this receptor was 141 ± 20 (DMSO) and 25 ± 7 (rif). The overall technical variance was small. For the nuclear receptor CAR and other very low expressed genes the variance was comparable to that of high expressed genes, but the impact on low expressed genes is much stronger. Similar expression levels were observed for the first three data points (filled circles, technical replicates of HH26) of the DMSO (blue) as well as Rifampicin (red) samples indicated by a almost horizontal lines in Figure 64.

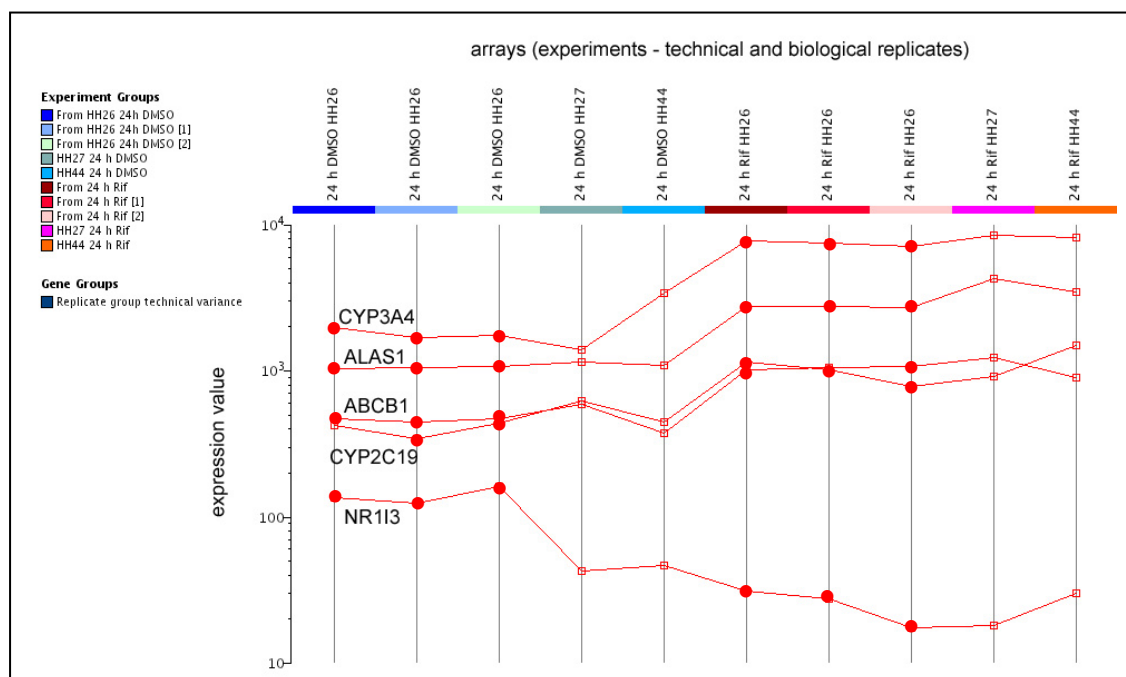


Figure 64: Parallel plot showing gene expression profiles for different biological and technical replicate arrays. The red line indicates gene expression levels of five genes (NR113, CYP2C19, ABCB1, ALAS1 and CAP3A4) which were investigated after 24 h of induction. On the y-axis the expression value is given and on the x-axis the different arrays are plotted. The blue arrays on the left were the controls (DMSO) and the first three (HH26) are technical replicates from the same biological sample prepared separately. The fourth and fifth blue labeled arrays (HH27 and HH44) are the second and third biological replicate of the control samples. For the Rifampicin arrays the same order is used (red). Filled circles indicate technical replicates.

The fourth and fifth data points represent the second and third biological replicate. The small open circles show a stronger variation compared with the expression level of the technical replicates. This difference is especially true for CYP3A4 and CAR (NR113). CYP3A4 shows an average expression in culture HH26 of 1826 ± 150 for DMSO. In culture HH27 the level was 1397 and in HH44 3461. After induction (Rif samples) the expression level were almost the same 7563 ± 300 for Rifampicin in culture HH26 8643 in HH27 and 8380 in HH44. For DMSO samples the CAR expression levels were 141 ± 20 (HH26), 43 and 47 (HH27 and HH44) and for the Rifampicin samples 25 ± 7 (HH26) 18 and 30 (HH27 and HH44). The levels for ALAS1 were very similar in all DMSO samples for the technical replicates as well as for the biological replicates. The induction of ALAS1 differed in the biological replicates. It was stronger in culture HH27 and HH44.

ABCB1 and CYP2C19 showed a variable expression pattern over different biological samples. The technical variability for ABCB1 was low. CYP2C19 showed a higher

technical as well as biological variability. A detailed overview of the technical variance (only for culture HH26) indicated by error bars and biological variance of the five example genes is given in Figure 65.

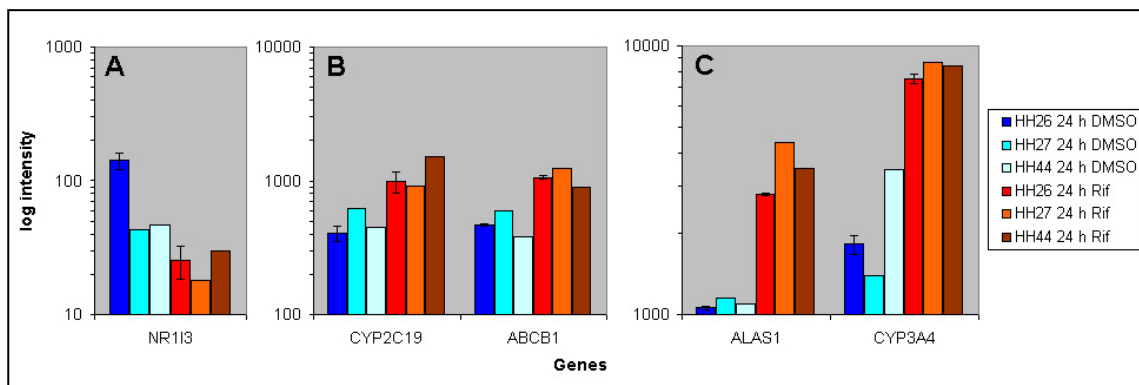


Figure 65: Technical and biological variance of different microarray hybridizations. The blue bars show the log intensity of arrays from control (DMSO) samples. The red bars give the intensities of Rifampicin treated samples. Each collection of bars represents one gene, shown on the x-axis. Each bar represents the signal intensity on the array for a specific sample. The first bar of each collection is an average of three technical replicates from cell culture HH26. The variance of this technical replicates is indicated by an error bar. The second bar shows the signal representative for the expression of the gene in culture HH27 and the third bar for culture HH44. In part A the low expressed nuclear receptor CAR is shown in B the medium expressed CYP2C19 and ABCB1 and in C the high expressed ALAS1 and CYP3A4.

Finally it could be concluded that the technical variance was low compared to the biological variability. For the high and medium expressed genes the variance was determined to be below 5 % (4.1 for DMSO and 2.7 for Rifampicin). However in case of very low expressed genes it could reach up to almost 30 %. The data for the group of low expressed genes had to be considered with special care. The biological variability resulted in different results especially for induced genes. The variance was between 4 and 70 % depending on the gene. In principle low expressed genes were burdened with a higher variance (e.g. the nuclear receptor CAR 72 %). The induction of high expressed genes, showed variable results in different samples. CYP3A4 as one example had a variance of 48.8 % in DMSO control samples, while the variance for the induced samples was only 6.9 %. The high expressed ALAS1 had a quite similar expression level in the control samples (DMSO) with a variance of 4.4 % over all three patients, while after induction with Rifampicin the signal intensity varied by 22.0 %.

These observations were also true for most of the other genes (data not shown). The technical variability was quite small over a large set of genes as indicated by the parallel plot in Figure 66.

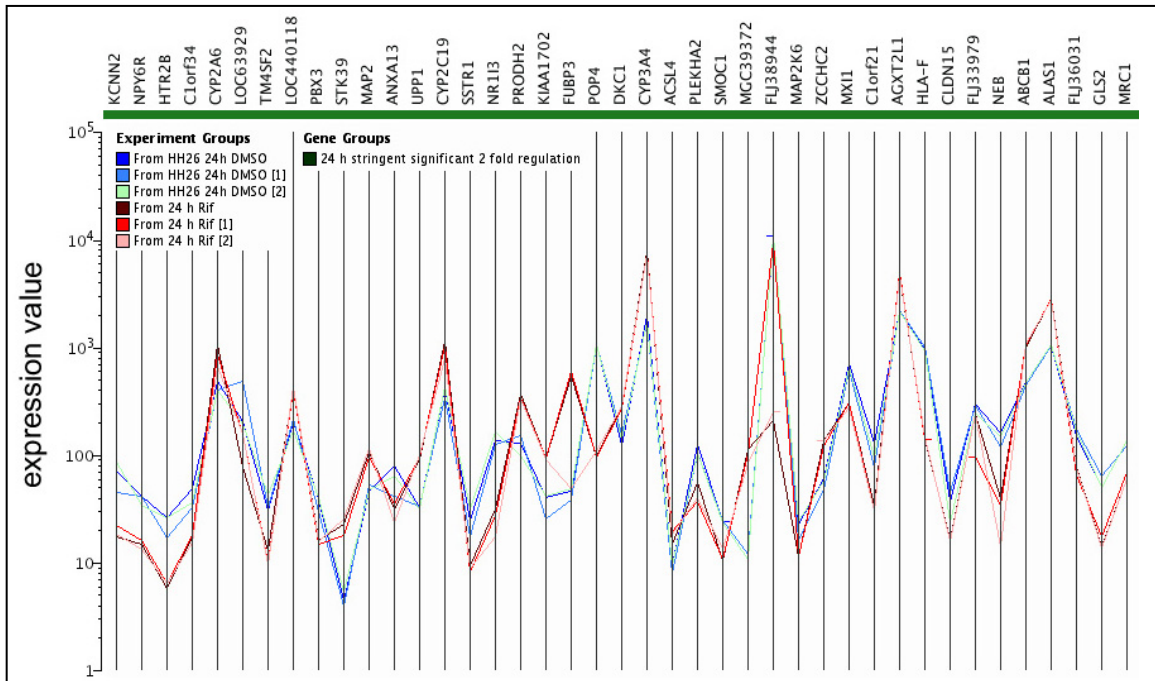


Figure 66: Parallel plot showing the variable expression of genes for the technical-replicates. The expression level of 41, 2-fold regulated genes was plotted in this graph. The intensity of the control samples is given in blue, while the intensity of the Rifampicin treated samples is shown as red curve. Overall a quite similar expression was observed between replicate samples.

The technical variability is plotted for the three replicates of culture HH26. The colour coded curves represent the expression levels of 41 different genes plotted on the x-axis. The expression level show a similar characteristic for the three control samples (blue) and the three induced samples (red). The overall expression levels were comparable between the three technical replicates. A lot of genes showed a similar behaviour for all three different patients (HH26, HH27 and HH44). In the following investigation of detecting differentially expressed genes the focus was set on the similar expressed genes in all three patients. If not mentioned otherwise the significance value for the observation in all three samples was set to $p = 0.05$.

3.8 Determination of Differentially Expressed Genes

After quantifying and correcting (normalization) of the signals, which correspond to the expression (mRNA) level of a gene in the cell, the search for differences between test

and reference samples was performed. This marker identification uses statistical tests for comparing different analysis groups (treatments, time, and patients), e.g. Rifampicin treated cells against control (DMSO) cells. One option to obtain marker (genes) for the difference between two groups is the n-fold regulation (proportions) analysis. This activity computes a score for each gene, measuring if a gene is equal to or higher than a given value e.g. 2-fold expressed in one sample compared to another.

3.8.1 Whole Genome Approach

Using this fold-change analysis to identify new candidates of genes which are up- or down-regulated in human hepatocytes after Rifampicin induction a whole genome approach was performed using HGU133 plus 2.0 GeneChips from Affymetrix. The experimental setup is presented in Figure 67.

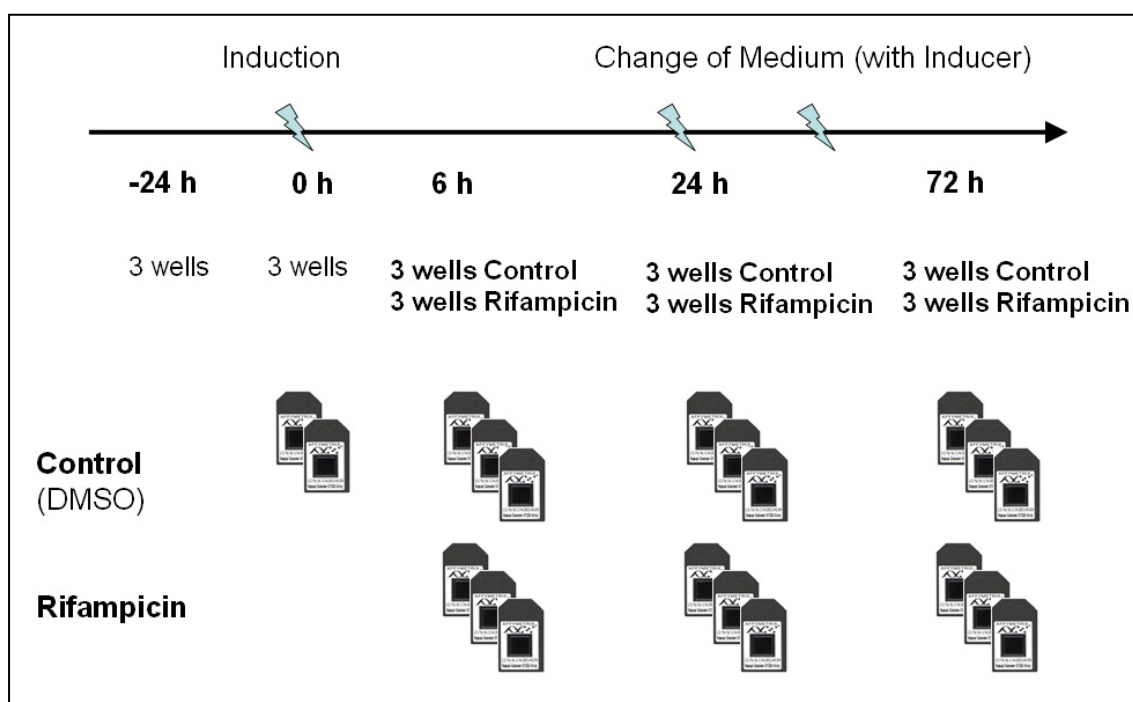


Figure 67: Experimental setup for whole genome expression profiling using hepatocyte cell cultures and Affymetrix arrays. The experimental design was based on time matched controls. Each measurement point (6, 24, 72 h) consisted of induced (Rifampicin) and not induced (DMSO) samples. For each sample three technical replicates were harvested and analysed. Every 24 h medium was changed (marked by an arrow on the timeline).

Human hepatocyte cell cultures were induced by Rifampicin at 0, 24 and 48 hours, indicated by the blue arrows. At five points in time (24 hours prior to induction "- 24 h", 0, 6, 24 and 72 h) samples were collected from three different wells of a six well plate. The RNA from each well (approximately 1.5 million cells) was extracted separately to maintain any possible differences in different wells. In our first approach using cell culture 26 (HH26), two replicates for the induction (0 h) and 6 h and three replicates for 24 and 72 h of induced- and control cells were analysed. The biological replicates consisting of cell culture 27 and 44 were analysed once on microarrays. To assess differentially expressed genes between induced and control samples, a fold change analysis followed by different statistical tests was performed.

3.8.1.1 Fold Change Analysis

Fold-change analysis was performed to identify genes, which had different mRNA concentrations in Rifampicin induced samples versus control samples. Intensities of two arrays to be compared are divided. The calculated ratio between these two intensities corresponds to the degree of difference (up- or down-regulation). Genes with a significant difference are likely to be involved in the regulation or directly in the detoxification of Rifampicin as they change their expression pattern in cells induced by Rifampicin. The cut off for expression ratios between Rifampicin samples and control samples was set to 2-fold. The number of valid values per array was set to 50 % meaning if a gene has less than 50 % of valid (non-missing) values on one of the arrays, all expression values of this gene in this experiment were ignored. For condensing of probe sets MAS5 algorithm (Schadt *et al.*, 2001) was used. A graphical representation of the expression values as scatter plot is given for each measurement point (6, 24 and 72 h) in the following graphs (Figure 68 - Figure 70).

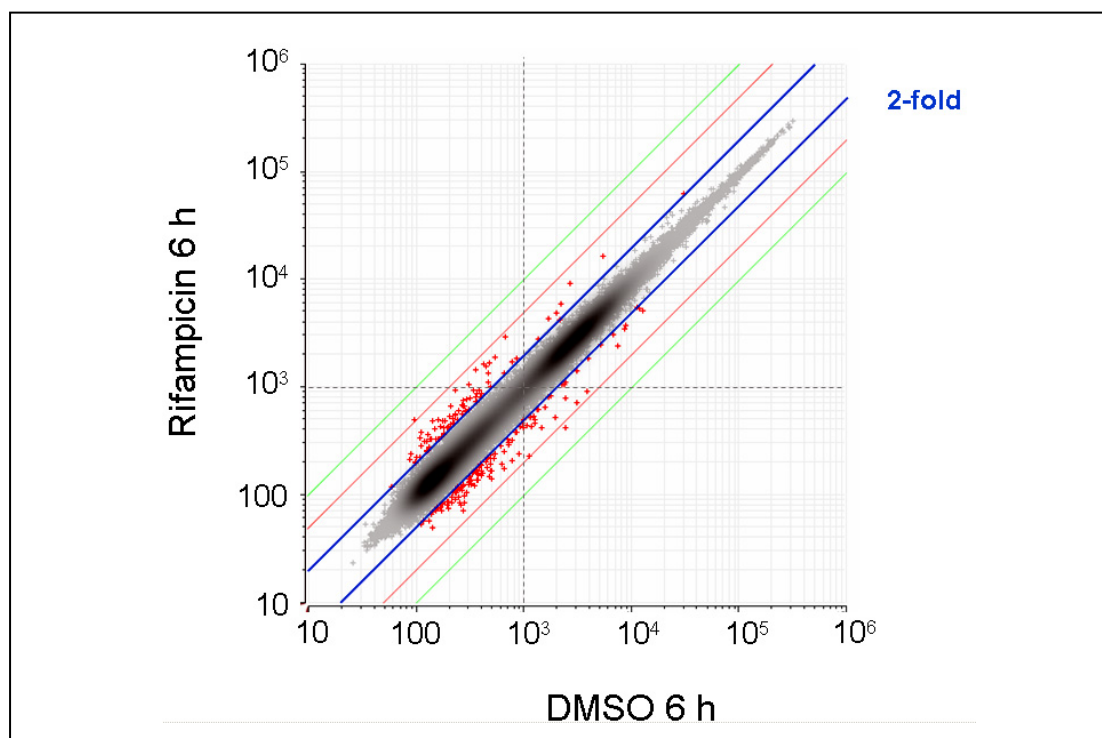


Figure 68: Scatter plot of logarithmic intensities for DMSO against Rifampicin arrays after 6 h of induction. Shown in red are the probes more than 2-fold differentially expressed.

The 2-fold analysis showed 265 regulated genes after 6 h (Figure 68), 475 after 24 h (Figure 69) and 345 after 72 h (Figure 70). After six hours (Figure 68) two genes showed a higher expression change than 5-fold (red lines). After 24 hours (Figure 69) of induction 7 genes showed a more than 5-fold induction and 6 genes a more than 5-fold down-regulation. The overall change of expression was much stronger compared to 6 hours. After 72 hours (Figure 70) the down-regulation of genes became more dominant compared to the up-regulation. Three genes showed a more the 5-fold induction while 14 genes were down-regulated. Some of these genes were down-regulated almost 10-fold.

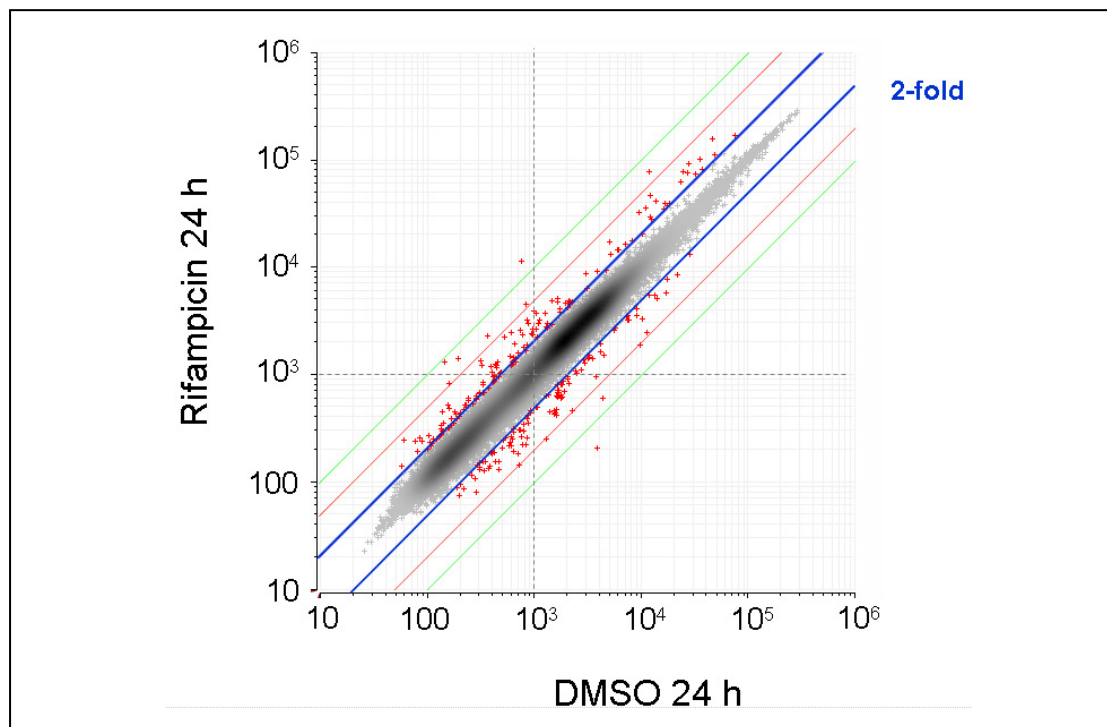


Figure 69: Scatter plot of logarithmic intensities for DMSO against Rifampicin arrays after 24 h of induction. Shown in red are the probes more than 2 fold differentially expressed.

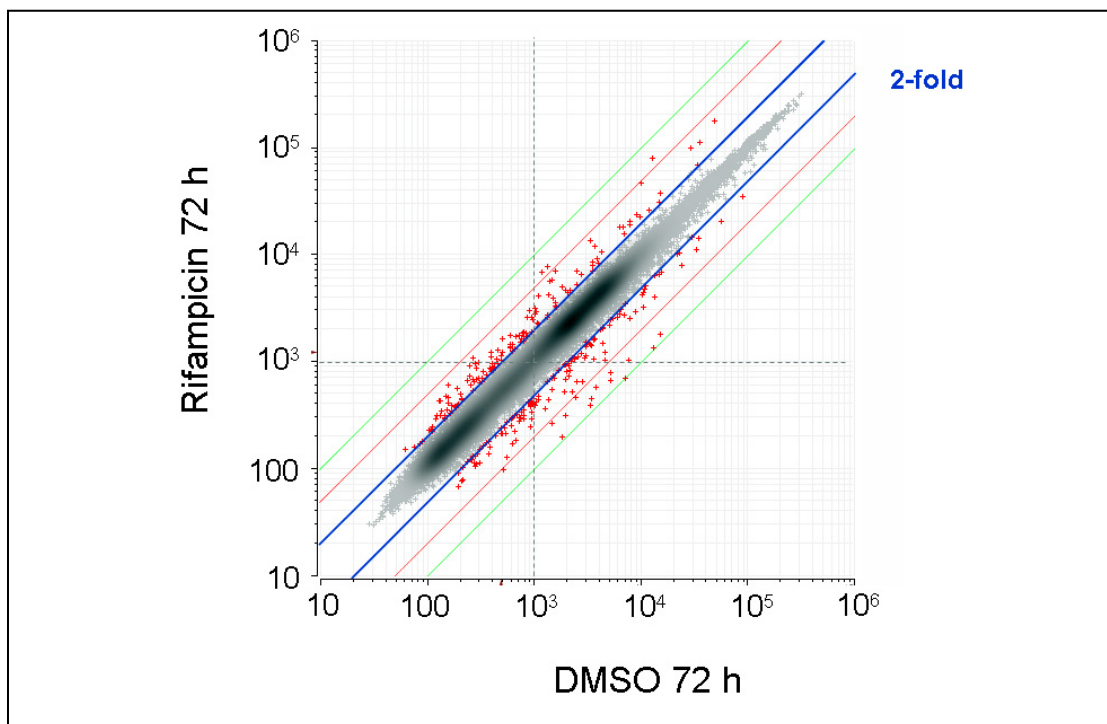


Figure 70: Scatter plot of logarithmic intensities for DMSO against Rifampicin arrays after 72 h of induction. Shown in red are the probes more than 2 fold differentially expressed.

As the n-fold regulation analyses does not produce p-values for the significance of an observation a subsequent statistical analysis was performed on the 2-fold regulated data set.

Several different statistical tests were applied to verify the total number of genes which showed at least 2-fold changed expression levels on different points in time between treated (Rifampicin) and control (DMSO) groups. Each test assigned a quantitative measure for the significance, the p-value. The cut-off p-value was set to $p = 0.05$. The test procedure computes a single number, the test statistic, from the input data. A theoretical distribution function gives the probability to observe a certain value for the test statistics, under certain assumptions about the absence of any effect (the null hypothesis). Finally, the p-value is obtained from integrating the distribution function over all values of the test statistics which are at least as far from its expectation value as the actually observed value. Therefore a low p-value stands for a high significance of the observation. The different tests performed were:

First the standard *t-test* which tests for the difference in means and assumes that the distribution of values is normal within each group and is identical for both groups. The second test was the *Welch test* which works as the t-test for the difference in means and assumes that the distribution of values is normal within each group. The Welch test in contrast does not require that the variance is identical in both groups. The third test applied was the *Kolmogorov-Smirnov test* which tests whether the distribution of values in the two groups differs. It makes no assumption on the distribution itself and is therefore very general. However, the disadvantage is that this test requires more samples than for example the t-test. Finally the *Wilcoxon test* which computes ranks from the values and tests whether the overlap is less than expected if the distributions in the two groups were the same. This is a non-parametric test. It makes no assumptions on the distribution of values within the groups. This test is also known as *Mann-Whitney test*. As there were three biological and three technical replicates analysed in this work, the t-test was the method of choice for further analysis.

Items were only considered for testing if they had enough valid values in each of the two groups. For all tests the number of items (experiments) per group should be at least three, better five or even more. With fewer experiments the reliability of the tests is questionable. The Wilcoxon and the Kolmogorov-Smirnov test work on ranks and

expect an even higher number of experiments per group so they were used and tested here as well, but the t-test based methods yield more reliable results.

A graphical overview of the workflow is given in Figure 71. On the left part of this overview the fold change operator is displayed. Fold changes are calculated based on the ratio of medians of the two experimental groups. The fold change threshold was set to 2 and only transcripts were used which show at least 50 % valid values per group. The resulting transcript group containing 265 regulated genes after 6 h, 475 after 24 h and 345 after 72 h was exported and stored ("more than 2 fold regulated"). In a sub-process "2 groups sample comparison workflow" the two experiment groups were analysed with the 2 group tests t-test, Wilcoxon, Welch, and Kolmogorov-Smirnov. The p-value cut-off was set to 0.05 and only transcripts were used which show at least 50 % valid values per group. The results from the 2 group tests were then combined ("significant ($p = 0.05$) and/or 2 fold regulation"). The intersection was also determined which means that only transcripts were taken which were significantly differentially expressed (p-values below 0.05) at the two group tests.

The four results from the 2 group tests, the union and the intersection were stored. The union of the 2 group tests was combined with the result from the fold change analysis resulting in genes having a minimum of 2-fold regulation or a p-value of maximum 0.05 in any of the 2-group tests. The intersection of the 2 group analysis was also intersected with the fold change result and this intersection was exported, resulting in genes differentially regulated in all of the four 2-group tests with a p-value of maximum 0.05 and having a minimum of 2-fold regulation ("string. signif. 2 fold regulation").

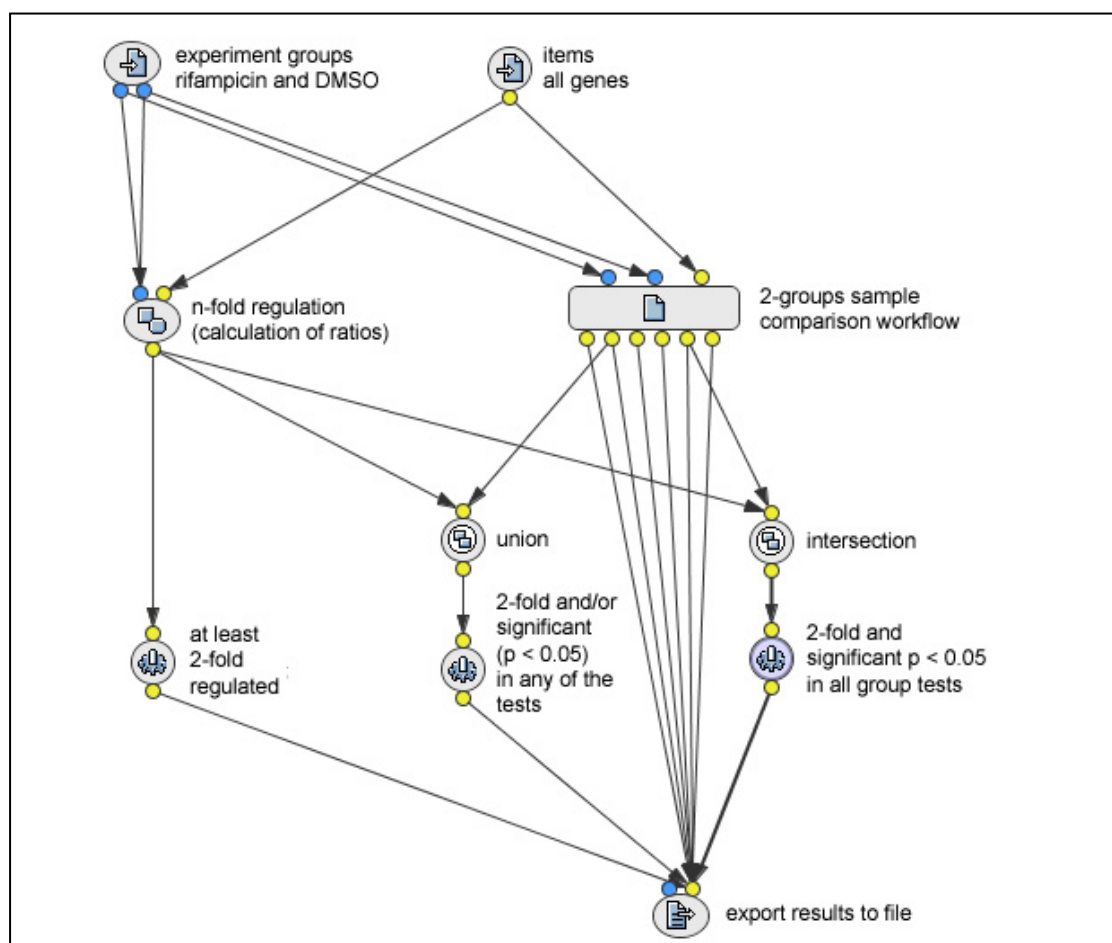


Figure 71: Graphical overview of the advanced 2 groups sample comparison workflow implemented in the Genedata Analyst software. 2 group comparison featuring 2 group statistical methods process and comparison with n-fold analysis performed in "Analyst".

At the end 7 groups were compiled: "more than 2 fold regulated" (from the fold change analysis calculated with ratio of medians), "t test $p \leq 0.05$ ", "Welch test $p \leq 0.05$ ", "Wilcoxon $p \leq 0.05$ ", "KS $p \leq 0.05$ ", "union of 2-groups tests results", "intersection of 2-groups tests results", "significant ($p = 0.05$) and/or 2-fold regulation" (union of fold change results and union of 2 group test results), and finally 'stringent significant 2-fold regulation' (intersection of fold change results and intersection of 2-group test results). In Table 39 a summary for the fold change analysis and statistical validation is given. The number of differentially expressed genes at the three investigated measurement points (6, 24 and 72 h) corresponding to each test is shown. A total number of 265 genes are more than 2-fold regulated after 6 h. The time point showing the highest number of 2-fold changed genes was 24 h (475 genes) while after 72 h the number of regulated genes decreased to 345. If significant regulated and 2-fold changed genes are combined the numbers increased to 866, 2215 and 1861 for 6, 24 and 72 h.

Table 39: This table shows the number of at least 2-fold and/or significantly regulated genes for all three time points based on different statistical tests.

regulation and statistical test	no. of diff. exp. probes		
	6 h	24 h	72 h
more than 2-fold regulated	265	475	345
significant ($p = 0.05$) and/or 2-fold regulation	866	2215	1861
T test $p = 0.05$	576	1914	1549
Welch test $p = 0.05$	438	1757	1382
Kolmogorov-Smirnov (KS) test $p = 0.05$	284	541	492
Wilcoxon test $p = 0.05$	284	1188	1121
string. signif. 2-fold regulation	17	77	49

Using different statistical tests, different numbers of regulated genes were obtained. However in all cases the largest number of differentially expressed genes was 24 h after induction with Rifampicin. If statistical significance from the different tests as well as a 2-fold cut off were applied to the data still 17 genes were identified for 6 h, 77 genes for 24 h and 49 genes for 72 h ($p = 0.05$). Figure 72 gives a graphical representation of the relationship between time, statistical test and the number of differentially expressed genes determined in Rifampicin versus control samples.

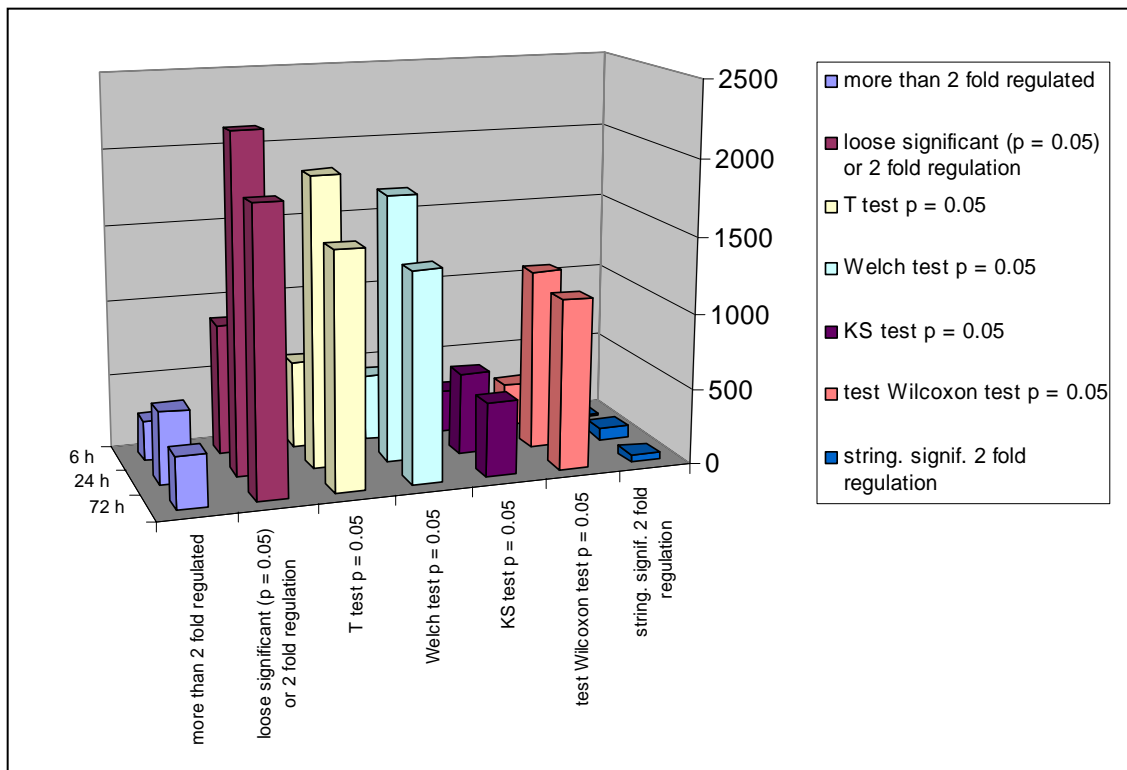


Figure 72: Graphical representation of the number of at least 2-fold and/or significantly regulated genes for all three time points based on different statistical tests. Dependent on the test (x-axis) the number of genes differs. Common patterns like time dependent change in expression levels is however visible in all tests in a similar way.

The t-test based methods (t-test, Welch) give higher and quite similar numbers which are more reliable compared to the ranking tests (KS and Wilcoxon). If all constraints are combined the number of items (genes) which fulfil all criteria (fold change and significance and valid values) dramatically decreases, however these are the most significant and interesting candidate genes. These genes identified by the fold-change analysis and proven to be similarly regulated in all patients by subsequent statistical tests are shown in Table 40 and Table 41.

Following these general observations a detailed time dependent analysis was performed. Using GC-RMA (Irizarry *et al.*, 2003) for condensing of different probes coding for one target gene, makes it dispensable to exclude expression values with low p-values (< 0.04) from the analysis. As GC-RMA creates a much lower noise for the low expressed genes compared to the previously used MAS5 algorithm this method was applied to increase the number of identified, especially low expressed detailed. The lower scattering, especially for these weak expressed genes can be observed in the log-log plots shown in Figure 74 - Figure 76. The quality filtering was disabled (max. p-value = 1). The raw data was median normalized using all probes to a fixed value of 10 followed by a pointwise division with the normalized DMSO group. The expression levels are visualized in a histogram of expression level (shown in Figure 73).

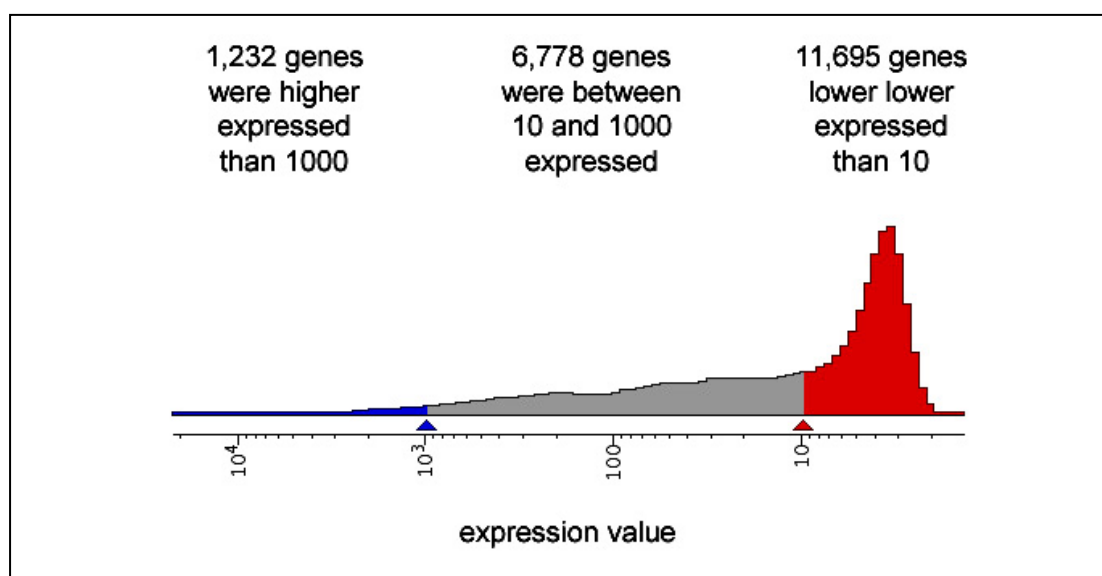


Figure 73: Histogram of expression levels determined with Affymetrix full genome arrays. Altogether 19,705 genes were taken into account. Distribution of high expressed genes (blue) medium expressed genes (grey) and low expressed genes (red) is visualized.

Table 40: Up-regulated genes found in all biological and technical replicates. Given for each gene are the corresponding probe-IDs from the Affymetrix array, a short gene description, the gene symbol and RefSeq ID number. The last five rows show some results from the performed analysis like regulation score, ratio of mean signals between Rifampicin and DMSO samples, the p-value ($p < 0.05$), the permutation q-value ($q < 0.05$) and the Benjamini Hochberg q-value ($q = 0.05$).

no.	Affymetrix probe ID	Gene Description	Gene Symbol	RefSeq Transcript	Regul. Score	Ratio of Means (Rif/DMSO)	p-value	Permutation Q-value	BH Q-value
1	243951_at	ATP-binding cassette, sub-family B memb. 1	ABCB1	NM_000927	2.43	2.43	9.34 E-06	0.0072	0.0143
2	221008_s_at	alanine-glyoxylate aminotransferase 2like	AGXT2L1	NM_031279	2.00	2.00	7.12 E-05	0.0074	0.0526
3	205633_s_at	aminolevulinate delta-synthase 1	ALAS1	NM_000688, NM_199166	2.62	2.62	3.19 E-04	0.0275	0.1211
4	215867_x_at	carbonic anhydrase XII	CA12	NM_001218, NM_206925	2.47	2.47	3.23 E-03	0.1350	0.3043
5	217133_x_at	cytochrome P450, family 2, subfam. B, polypeptide 6	CYP2B6	NM_000767	3.44	3.44	1.56 E-03	0.0762	0.2259
6	216058_s_at	cytochrome P450, family 2, subfam. C, polypeptide 19	CYP2C19	NM_000769	2.20	2.20	1.10 E-03	0.0575	0.1934
7	208147_s_at	cytochrome P450, family 2, subfam. C, polypeptide 8	CYP2C8	NM_000770, NM_030878	2.72	2.72	4.13 E-03	0.1484	0.3306
8	208367_x_at	cytochrome P450, family 3, subfam. A, polypeptide 4	CYP3A4	NM_017460	4.46	4.46	5.79 E-04	0.0424	0.1481
9	211442_x_at	cytochrome P450, family 3, subfam. A, polypeptide 43	CYP3A43	NM_022820, NM_057095, NM_057096	2.25	2.25	4.73 E-03	0.1586	0.3421
10	211843_x_at	cytochrome P450, family 3, subfam. A, polypeptide 7	CYP3A7	NM_000765	2.52	2.52	1.10 E-03	0.0575	0.1934
11	221577_x_at	growth differentiation factor 15	GDF15	NM_004864	2.09	2.09	8.71 E-03	0.2116	0.4070
12	243614_s_at	proline dehydrogenase/oxidase 2	PRODH2	NM_021232	2.19	2.19	5.68 E-03	0.1745	0.3706
13	209723_at	serine or cysteine proteinase inhibitor, clade B member 9	SERPINB9	NM_004155	2.55	2.55	4.72 E-08	0.0072	0.0003
14	212811_x_at	solute carrier family 1, member 4	SLC1A4	NM_003038	2.65	2.65	6.10 E-04	0.0434	0.1522
15	212295_s_at	solute carrier family 7 cationic amino acid trans-porter member 1	SLC7A1	NM_003045	2.00	2.00	6.53 E-03	0.1909	0.3851
16	202786_at	serine threonine kinase 39	STK39	NM_013233	2.84	2.84	9.88 E-04	0.0524	0.1809
17	216623_x_at	trinucleotide repeat containing	TNRC9	XM_049037	3.87	3.87	3.80 E-05	0.0074	0.0421

Table 41: Down-regulated genes found in all biological and technical replicates. Given are the corresponding probe-IDs from the Affymetrix array, a short gene description, the gene symbol and RefSeq ID number. The last five rows show some results from the performed analysis like regulation score, ratio of mean signals between Rifampicin and DMSO samples, the p-value ($p < 0.05$), the permutation q-value ($q < 0.05$) and the Benjamini Hochberg q-value ($q = 0.05$)

No.	Affymetrix probe ID	Gene Description	Gene Symbol	RefSeq Transcript	Regul. Score	Ratio of Means (Rif/DMSO)	p-value	Permutation Q-value	BH Q-value
1	209612_s_at	alcohol dehydrogenase IB (class I), beta polypeptide	ADH1B	NM_000668	9.96	0.10	2.74 E-04	0.0255	0.1092
2	231678_s_at	alcohol dehydrogenase 4 (class II), pi polypeptide	ADH4	NM_000670	2.40	0.42	2.76 E-03	0.1220	0.2858
3	214261_s_at	alcohol dehydrogenase 6	ADH6	NM_000672	2.02	0.49	5.30 E-03	0.1720	0.3641
4	204685_s_at	ATPase, Ca ⁺⁺ transporting, plasma membrane 2	ATP2B2	NM_001683	2.22	0.45	6.58 E-03	0.1909	0.3851
5	1431_at	cytochrome P450, family 2, subfamily E, polypeptide 1	CYP2E1	NM_000773	3.34	0.30	3.48 E-04	0.0275	0.1211
6	233979_s_at	espin	ESPN	NM_031475	2.20	0.45	5.49 E-03	0.1735	0.3675
7	1564706_s_at	glutaminase 2 (liver, mitochondrial)	GA, GLS2	NM_013267, NM_138566	2.16	0.46	4.67 E-03	0.1586	0.3405
8	238752_at	Glycosylphosphatidylinositol specific phospho-lipase D1	GPLD1	NM_001503, NM_177483	2.43	0.41	1.68 E-03	0.0778	0.2299
9	214621_at	glycogen synthase 2 (liver)	GYS2	NM_021957	2.41	0.42	9.48 E-03	0.2252	0.4268
10	231156_at	Hydroxyacid oxidase 2 (long chain)	HAO2	NM_016527	3.21	0.31	6.27 E-03	0.1877	0.3844
11	204438_at	mannose receptor, C type 1	MRC1	NM_002438	2.08	0.48	8.82 E-07	0.0072	0.0025
12	220554_at	solute carrier family 22 (organic anion transporter), member 7	SLC22A7	NM_006672, NM_153320	2.19	0.46	3.12 E-03	0.1350	0.3030
13	226625_at	Transforming growth factor, beta receptor III (betaglycan, 300kDa)	TGFBR3	NM_003243	2.41	0.42	1.72 E-03	0.0780	0.2301

1,232 genes had an expression level > 1000 . 6,778 genes were expressed in the range $1000 < x < 10$. 11,695 genes were low expressed genes with an expression value below ten. Altogether 19,705 genes were included in the subsequent analysis.

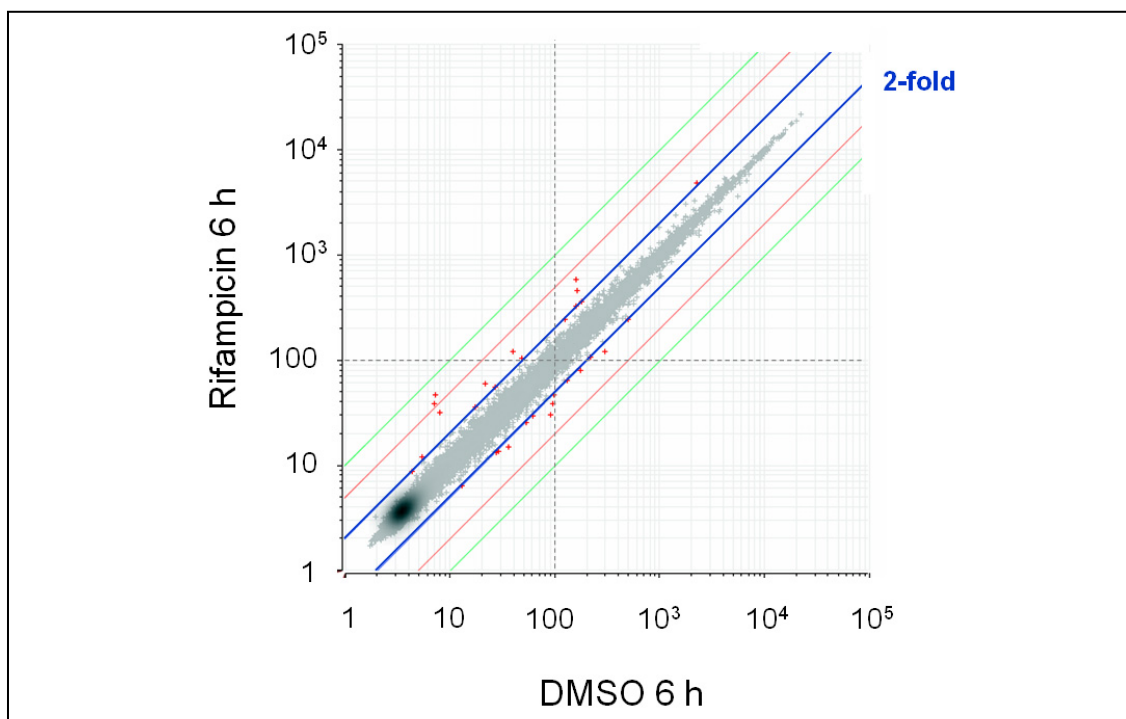


Figure 74: Scatter plot of logarithmic intensities for DMSO against Rifampicin arrays after 6 h of induction. Shown in red are the probes more than 2 fold differentially expressed.

The detected differentially expressed genes for each time point are given in Table 42 - Table 47. These tables present only the genes showing the highest fold change in the accordant time point. A summary of all more than 2-fold differentially expressed genes is given in the appendix (Table 58 - Table 63).

Cytochrome P450, family 26, subfamily A, polypeptide 1 (CYP26A1) showed the highest up-regulation after 6 h of induction with Rifampicin. The up-regulation was measured to be 6.5-fold after 6 h. After 24 and 72 h the corresponding gene was not detected to be over-expressed anymore. This gene encodes a member of the cytochrome P450 superfamily of enzymes which are monooxygenases catalyzing reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.

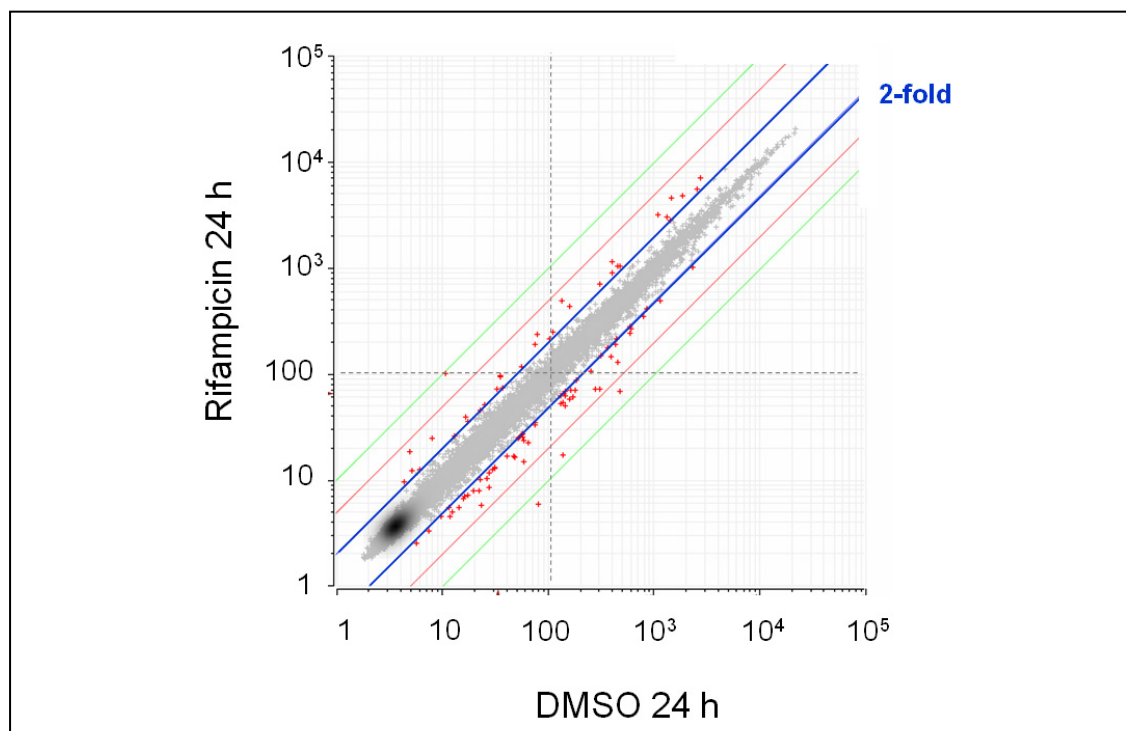


Figure 75: Scatter plot of logarithmic intensities for DMSO against Rifampicin arrays after 24 h of induction. Shown in red are the probes more than 2 fold differentially expressed.

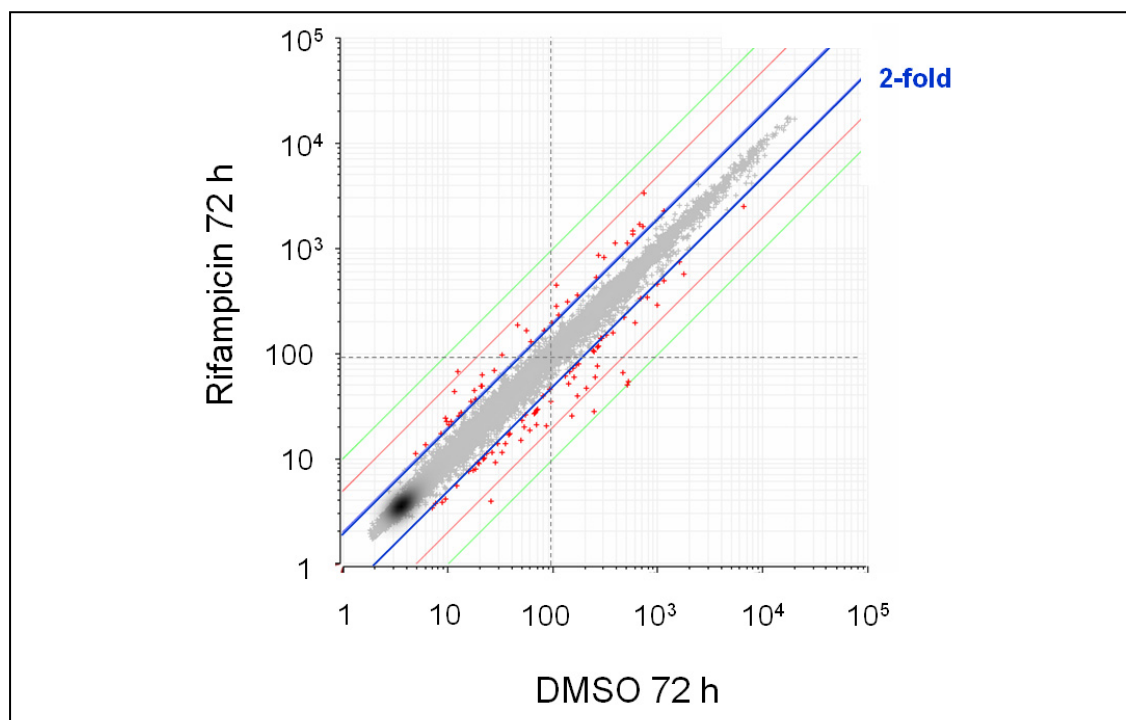


Figure 76: Scatter plot of logarithmic intensities for DMSO against Rifampicin arrays after 72 h of induction. Shown in red are the probes more than 2 fold differentially expressed.

This endoplasmic reticulum protein acts on retinoids, including all-trans-retinoic acid (RA), with both 4-hydroxylation and 18-hydroxylation activities. It regulates the cellular level of retinoic acid which is involved in regulation of gene expression in both embryonic and adult tissues. The second most up-regulated (3.7 fold) gene was thyroid hormone responsive SPOT14; thyroid hormone-inducible hepatic protein (THRSP).

Table 42: Five up-regulated annotated genes in primary hepatocytes showing the highest fold change after 6 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 6 h of induction. A complete list of 2-fold up-regulated genes after 6 h of induction can be found in the appendix.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	CYP26A1	cytochrome P450, family 26, subfamily A, polypeptide 1	6.5	1.0	1.0
2	THRSP	thyroid hormone responsive (SPOT14 homolog, rat)	3.7	1.7	1.0
3	TNRC9	trinucleotide repeat containing 9	3.1	3.0	4.1
4	FASN	fatty acid synthase	2.8	1.4	0.5
5	SH3TC1	SH3 domain and tetratricopeptide repeats 1	2.8	1.7	1.1

Table 43: Five down-regulated annotated genes in primary hepatocytes showing the highest fold change after 6 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 6 h of induction. A complete list of 2-fold down-regulated genes after 6 h of induction can be found in the appendix.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	ADH1B	alcohol dehydrogenase (class I), beta polypeptide	0.3	0.1	0.1
2	SLC38A4	solute carrier family 38, member 4	0.4	0.5	0.5
3	GLS2	glutaminase 2 (liver, mitochondrial)	0.5	0.3	0.5
4	SLC6A1	solute carrier family 6 (neurotransmitter transporter, GABA), member 1	0.5	0.4	0.9
5	GPR64	G protein-coupled receptor 64	0.5	1.0	0.1

The protein encoded by this gene is similar to the gene product of S14, a rat gene whose expression is limited to liver and adipose tissue and is controlled by nutritional and hormonal factors. This gene has been shown to be expressed in liver and adipocytes (the cells specialized in storing energy as fat), particularly in lipomatous modules and is suggested to play a role in controlling lipid metabolism. Over time the expression of this gene goes down to a not over-expressed state. After 24 h the expression is 1.7-fold and after 72 h no differentially expression is observed. The gene showing the third highest up-regulation after 6 h was TOX high mobility group box family member 3 (TOX3, TNRC9). This gene was up-regulated more than 2-fold for all three investigated time points. After 6 h it showed a 3.1-fold up-regulation. It was provisional classified to have a functional role in DNA binding but a specific function was not assigned so far. The fatty acid synthase (FASN) and SH3 domain and tetratricopeptide repeats 1 (SH3TC1) are both up-regulated 2.8-fold after 6 h. The fatty acid synthase is a multifunctional protein. Its main function is to catalyze the synthesis of palmitate from acetyl-CoA and malonyl-CoA, in the presence of NADPH, into long-chain saturated fatty acids. In some cancer cell lines, this protein has been found to be fused with estrogen receptor-alpha (ER-alpha), in which the N-terminus of FAS is fused in-frame with the C-terminus of ER-alpha. Over time the expression declines from 2.8-fold over-expressed after 6 h to 1.4 over-expressed after 24 h and after 72 h a 2-fold down-regulation was observed. The SH3TC1 has so far no functional annotation however the expression was also declining from 2.8- to 1.7- to 1.1-fold over-expression.

The gene showing the highest down-regulation after 6 h was alcohol dehydrogenase IB (class I), beta polypeptide (ADH1B, ADH2) with a 3.3-fold down-regulation. The down-regulation continued over time and after 24 und 72 h this gene was 10-fold down-regulated. The protein encoded by this gene is a member of the alcohol dehydrogenase family. Members of this enzyme family metabolize a wide variety of substrates, including ethanol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products. This encoded protein, consisting of several homo- and heterodimers of alpha, beta, and gamma subunits, exhibits high activity for ethanol oxidation and plays a major role in ethanol catabolism. Three genes encoding alpha, beta and gamma subunits are tandemly organized in a genomic segment as a gene cluster, however only the gene coding for the beta polypeptide was so strongly down-

regulated. Another gene showing a down-regulation in all time points was solute carrier family 38, member 4 (SLC38A4) which showed a 2.5-fold down-regulation after 6 h and 2-fold down-regulation for the two later time points. SLC38A4 is found predominantly in liver and transports both cationic and neutral amino acids. The transport of cationic amino acids by SLC38A4 is Na⁺ and pH independent, while the transport of neutral amino acids is Na⁺ and pH dependent (Hatanaka et al., 2001). Further 2-fold down-regulated genes after 6 h of Rifampicin induction were *gls2*, *slc6A1* and *gpr64*. Glutaminase 2 - liver, mitochondrial (GLS2) is a phosphate-activated glutaminase that catalyzes the hydrolysis of glutamine to stoichiometric amounts of glutamate and ammonia. Solute carrier family 6 (neurotransmitter transporter, GABA), member 1 (SLC6A1) has so far a preliminary functional annotation which suggests a role for development of cognitive abilities, associative learning and new object recognition as a neurotransmitter transporter. G protein-coupled receptor 64 (GPR64) plays a role in signal transduction however the explicit function is not clear so far. A role in male fertility and sperm maturation is discussed.

After 24 hours the effect of the induction with Rifampicin had a stronger impact on the change of expression level. The ten genes showing the highest change in gene expression after 24 h of induction are summarized in Table 44 and Table 45.

The strongest up-regulation after 24 h was observed for the serine threonine kinase 39 (STE20/SPS1 homolog, yeast) (STK39) with a factor of 3.8-fold. This gene encodes a serine/threonine kinase that is thought to function in the cellular stress response pathway. The kinase is activated in response to hypotonic stress, leading to phosphorylation of several cation-chloride-coupled co-transporters. The catalytically active kinase specifically activates the p38 MAP kinase pathway, and its interaction with p38 decreases upon cellular stress, suggesting that this kinase may serve as an intermediate in the response to cellular stress. It was already up-regulated slightly after 6 h with a factor of 1.5-fold and is still higher expressed in Rifampicin cells compared to control (DMSO) cells after 72 h by a factor of 2.4-fold. Cytochrome P450, family 2, subfamily B, polypeptide 6 (CYP2B6), another member of the cytochrome P450 superfamily, was the second most up-regulated (3.7-fold) protein after 24 h. This protein localizes to the endoplasmic reticulum and its expression is known to be induced by phenobarbital. This observation suggests that this is also true for

Rifampicin. The enzyme is so far also known to metabolize some xenobiotics, such as the anti-cancer drugs cyclophosphamide and ifosfamide.

Table 44: Ten up-regulated annotated genes in primary hepatocytes showing the highest fold change after 24 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 24 h of induction. A complete list of more than 2-fold up-regulated genes after 24 h of induction can be found in the appendix.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	STK39	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)	1.5	3.8	2.4
2	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	2.0	3.7	3.0
3	MT1F	metallothionein 1F (functional)	1.2	3.2	0.9
4	TNRC9	trinucleotide repeat containing 9	3.1	3.0	4.1
5	ALAS1	5-aminolevulinate synthase	2.2	2.9	2.6
6	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	1.6	2.9	2.9
7	PRODH2	proline dehydrogenase (oxidase) 2	1.6	2.8	2.0
8	SEC14L4	SEC14-like 4 (<i>S. cerevisiae</i>)	1.9	2.8	2.0
9	UPP1	uridine phosphorylase 1	1.9	2.7	1.3
10	MT1X	metallothionein 1X	1.0	2.6	0.8

Other important up-regulated genes after 24 h were: *mt1f*, *alas1*, *cyp3A4* and *mt1x*. Metallothionein 1F (MT1F) showed a 3.2-fold up-regulation and metallothionein 1X (MT1X) a 2.6-fold up-regulation. In the context of liver cancer there was an observation that MT1F shows down-regulated expression which supports the inhibited function of MT1F in cancer growth. In general metallothionein proteins have metal ion binding functionality and therefore protect against metal toxicity. Further is discussed whether this family of enzymes also play a role in protection against oxidative stress. The strong induction was only observed after 24 h. After 6 h of induction no change in expression was observed and after 72 h of induction both genes were not differentially expressed. Aminolevulinate, delta-, synthase 1 (ALAS1) was up-regulated 2.9-fold after 24 h of induction. The gene expression of this enzyme is regulated by Hepatic nuclear

factor 3 and nuclear factor 1. The enzyme is involved in the porphyrin metabolism hence delivering heme for the active centres of the cytochrome P450 enzymes.

Another member of this CYP-family (CYP3A4) was found to be 2.9-fold up-regulated after 24 h of induction. Cytochrome P450, family 3, subfamily A, polypeptide 4 localizes to the endoplasmic reticulum and its expression is induced by glucocorticoids and some pharmacological agents like Rifampicin. This enzyme is involved in the metabolism of approximately half the drugs which are used today, including acetaminophen, codeine, cyclosporin A, diazepam and erythromycin. This enzyme is therefore of high importance and main target of our investigations.

Furthermore TNRC9 (3.0-fold), PRODH2 (2.8-fold), SEC14L4 (2.8-fold) and UPP1 (2.7-fold) were detected to be up-regulated.

After 24 h of induction, three genes showed a high down-regulation of 10-fold. These strongly down-regulated genes in Rifampicin cells compared to DMSO cells were *cyp7A1*, *adh1a* and *adh1B*. Additionally *adh1c* (5-fold) and *cyp2e1* (3.3-fold) also involved in the metabolism of alcohol were also detected to be down-regulated. Cytochrome P450, family 7, subfamily A, polypeptide 1 (CYP7A1) is an endoplasmic reticulum membrane protein which catalyzes the first reaction in the cholesterol catabolic pathway in the liver. In this catabolic pathway cholesterol is converted to bile acids. This reaction is the rate limiting step and the major site of regulation of bile acid synthesis, which is the primary mechanism for the removal of cholesterol from the body. CYP7A1 was detected to be down-regulated after 24 h as well as 72 h after induction. Also for 24 and 72 h the genes coding for the alpha-, beta- and gamma-polypeptide of the alcohol dehydrogenase were detected to be strongly down-regulated. CYP2E1 shown to be down-regulated 3.3-fold metabolizes both endogenous substrates, such as ethanol, acetone, and acetal, as well as exogenous. Due to its many substrates, this enzyme may be involved in such varied processes as gluconeogenesis, hepatic cirrhosis, diabetes, and cancer. Further 3-fold down-regulated genes were *htr2b*, *gls2*, *ifit1*, *aspa* and *amy2a*.

Table 45: Ten down-regulated annotated genes in primary hepatocytes showing the highest fold change after 24 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 24 h of induction. A complete list of 2-fold down-regulated genes after 24 h of induction can be found in the appendix.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	CYP7A1	cytochrome P450 7A1	0.6	0.1	0.2
2	ADH1A	alcohol dehydrogenase (class I), alpha polypeptide	0.7	0.1	0.2
3	ADH1B	alcohol dehydrogenase (class I), beta polypeptide	0.3	0.1	0.1
4	ADH1C	alcohol dehydrogenase 1C (class 1), gamma polypeptide	0.6	0.2	0.2
5	HTR2B	5-hydroxytryptamine (serotonin) receptor 2B	0.5	0.3	0.7
6	GLS2	glutaminase 2 (liver, mitochondrial)	0.5	0.3	0.5
7	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	0.4	0.3	0.8
8	CYP2E1	cytochrome P450 2E1	0.7	0.3	0.2
9	ASPA	ASPARTATE AMMONIA-LYASE	0.6	0.3	1.0
10	AMY2A	amylase, alpha 2A; pancreatic	0.7	0.4	0.5

As detected for the previous time point (24 h) TNRC9 (4.1-fold), CYP2B6 (3.0-fold), CYP3A4 (2.9-fold) and ALAS1 (2.6-fold) were up-regulated after 72 h. In addition CYP2C8, NPPB, CCL2, TLR1, CRP and SLC13A5 were higher expressed in Rifampicin treated cells compared to DMSO cells. Cytochrome P450, family 2, subfamily C, polypeptide 8 (CYP2C8) showed the highest up-regulation after 72 h with a factor of 4.7-fold. The enzyme is known to metabolize many xenobiotics, including the anticonvulsive drug mephenytoin, benzo[a]pyrene, 7-ethoxycoumarin, and the anti-cancer drug taxol. Natriuretic peptide precursor B (NPPB) is a member of the natriuretic peptide family and encodes a secreted protein which functions as a cardiac hormone. The protein undergoes two cleavage events, one within the cell and a second after secretion into the blood. The protein's biological actions include natriuresis, diuresis, vasorelaxation, inhibition of renin and aldosterone secretion and a key role in

cardiovascular homeostasis. A high concentration of this protein in the bloodstream is indicative of heart failure. Chemokine (C-C motif) ligand 2 (CCL2) is one of several cytokine genes clustered on the q-arm of chromosome 17. Cytokines are a family of secreted proteins involved in immune-regulatory and inflammatory processes. The protein encoded by this gene is structurally related to the CXC subfamily of cytokines. This cytokine displays chemotactic activity for monocytes and basophils but not for neutrophils or eosinophils. It has been implicated in the pathogenesis of diseases characterized by monocytic infiltrates, like psoriasis, rheumatoid arthritis and atherosclerosis. It binds to chemokine receptors CCR2 and CCR4. Toll-like receptor 1 (TLR1) is a member of the Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of immunity. TLRs are highly conserved from *Drosophila* to humans and share structural and functional similarities. They recognize pathogen-associated molecular patterns (PAMPs) that are expressed on infectious agents, and mediate the production of cytokines necessary for the development of effective immunity. The various TLRs exhibit different patterns of expression. TLR1 gene is ubiquitously expressed, and at higher levels than other TLR genes.

C-reactive protein (CRP) is an acute phase protein produced by the liver. Its levels rise dramatically during inflammatory processes occurring in the body. This increment is due to a rise in the plasma concentration of IL-6, which is produced in macrophages, endothelial cells and T-cells. This increase is used as a marker of inflammation. Measuring C-reactive protein values can prove to be useful in determining disease progress or the effectiveness of treatments.

Solute carrier family 13 (sodium-dependent citrate transporter), member 5 (SLC13A5, NACT) is a tricarboxylate plasma transporter with a preference for citrate (Na⁺-coupled citrate uptake), expressed in the sinusoidal membrane of hepatocytes.

As shown for the previous time points the following genes were down-regulated as well after 72 h: *adh1a*, *adh1b*, *adh1c*, *gpr64*, *cyp7a1* and *cyp2e1*. In addition to those genes *hao2* (10-fold), *gys2* (5-fold), *slc22a7* (2.5-fold) and *sult1b1* (2.5-fold) were found to be strongly down-regulated. Hydroxyacid oxidase 2 (HAO2) has 2-hydroxyacid oxidase activity, is localized in peroxisomes and expressed predominantly in liver and kidney. Glycogen synthase 2 - liver (GYS2, UDP-glucose-glycogen glucosyltransferase) is a

glycosyltransferase enzyme that catalyses the transfer of glucose residues one by one into a polymeric chain for storage as glycogen. Solute carrier family 22 (organic anion transporter), member 7 (SLC22A7) is involved in the sodium-independent transport and excretion of organic anions, some of which are potentially toxic. Sulfotransferase family, cytosolic, 1B, member 1 (SULT1B1) catalyze the sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds. These cytosolic enzymes are different in their tissue distributions and substrate specificities.

Table 46: Ten up-regulated annotated genes in primary hepatocytes showing the highest fold change after 72 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 72 h of induction. A complete list of more than 2-fold up-regulated genes after 72 h of induction can be found in the appendix.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8	1.7	2.2	4.7
2	TNRC9	trinucleotide repeat containing 9	3.1	3.0	4.1
3	NPPB	natriuretic peptide precursor B	1.5	1.6	4.0
4	CCL2	chemokine (C-C motif) ligand 2	1.3	2.6	3.2
5	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	2.0	3.7	3.0
6	TLR1	Toll-like receptor 1	1.5	1.4	3.0
7	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	1.6	2.9	2.9
8	CRP	C-reactive protein	1.2	1.4	2.7
9	SLC13A5	solute carrier family 13 (sodium-dependent citrate transporter), member 5	1.2	1.5	2.6
10	ALAS1	5-aminolevulinate synthase	2.2	2.9	2.6

Table 47: Ten down-regulated annotated genes in primary hepatocytes showing the highest fold change after 72 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 72 h of induction. A complete list of 2-fold down-regulated genes after 72 h of induction can be found in the appendix.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	ADH1B	alcohol dehydrogenase (class I), beta polypeptide	0.3	0.1	0.1
2	GPR64	G protein-coupled receptor 64	0.5	1.0	0.1
3	HAO2	hydroxyacid oxidase 2 (long chain)	0.6	0.4	0.1
4	CYP7A1	cytochrome P450 7A1	0.6	0.1	0.2
5	ADH1A	alcohol dehydrogenase (class I), alpha polypeptide	0.7	0.1	0.2
6	GYS2	glycogen synthase 2 (liver)	0.7	0.5	0.2
7	CYP2E1	cytochrome P450 2E1	0.7	0.3	0.2
8	ADH1C	alcohol dehydrogenase 1C (class 1), gamma polypeptide	0.6	0.2	0.2
9	SLC22A7	solute carrier family 22 (organic anion transp.), member 7	0.9	0.5	0.4
10	SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	0.9	0.6	0.4

While after 6 h only 30 genes were ≥ 2 -fold differentially expressed, the number of differentially expressed genes increased strongly for the later time points. After 24 h of induction 142 genes altered their expression ≥ 2 -fold. After 72 h 137 genes were at least 2-fold or higher differentially expressed.

Seven genes changed their expression profile in a similar way in all time points. Two of these seven genes were not functional annotated. One gene was classified as coding for a hypothetical protein MGC39372. The second not functional annotated probe was LOC343071 which is said to be similar to hypothetical protein DJ845O24.1. The remaining five genes were: serine threonine kinase 39 (STK39), cytochrome P450, family 2, subfamily B, polypeptide 6 (CYP2B6), trinucleotide repeat containing 9 (TNRC9), 5-aminolevulinate synthase (ALAS1) and alcohol dehydrogenase (class I), beta polypeptide (ADH1B). Except for ADH1B which was detected to be down-regulated in all time points, the other genes similarly expressed were up-regulated. These observations are summarized in two Venn diagrams shown in Figure 77.

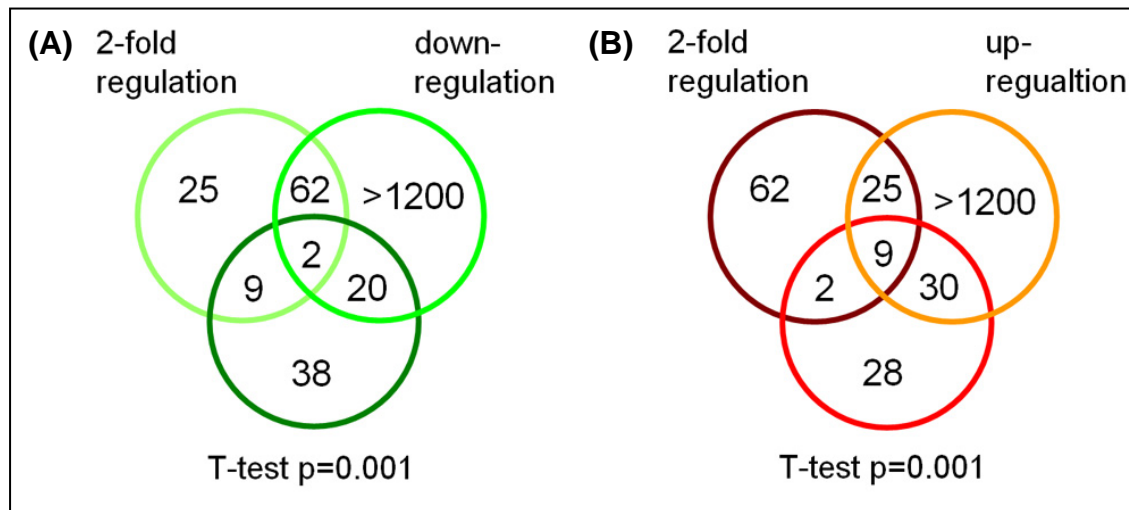


Figure 77: Venn diagrams showing the number of 2-fold, up- and down-regulated genes as well as significantly expressed genes in all patients after 24 h of Rifampicin induction. Part (A) shows the down-regulated genes. 98 genes were 2-fold regulated, more than 1200 were down-regulated and 69 genes were highly significant expressed in all cell cultures. 62 genes showed a 2-fold or higher down-regulation and 2 of this genes were further highly significant expressed in all samples. From the 98 2-fold differentially expressed genes 25 were up-regulated and nine were significantly up-regulated in all three patients as displayed in part (B).

A helpful visualization tool for microarray data is the so called volcano plot. It is a kind of scatter plot of genes in which p-values (y-axis) are plotted against n-fold change of expression (x-axis). The lower the values on the y-axis the more significant is the observation in the three biological replicates. Genes found on the right side of the plot are up-regulated and genes on the left side are down-regulated. The distance from the middle (1.0) corresponds to the fold up- or down-regulation. A gene in the top right area is significantly up-regulated. If a gene is on the left lower part it is down-regulated but this observation is not very significant. Concerning change of expression the most interesting genes are in the left and right upper corners of such a plot.

Figure 78 shows a volcano plot for the expression profiles of three primary human hepatocyte cell cultures after 6 hours of Rifampicin treatment. The fold change is determined by comparing induced versus control samples. Seven genes show a more than 2-fold significant differential expression. These genes are indicated by red dots and given gene name.

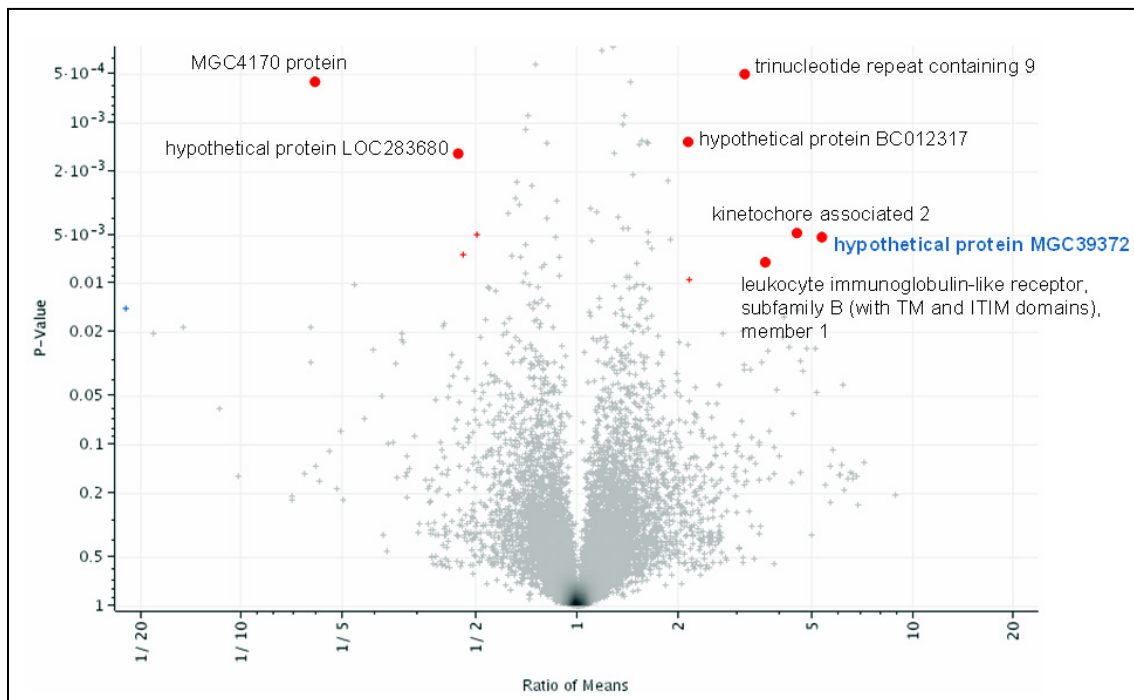


Figure 78: Volcano plot of expression data from samples treated 6 h with Rifampicin. Ten genes shown in red matching the criteria (2 fold and $p = 0.01$). The seven most significant (low p -value) are given with their associated name/function.

Four of these seven significantly differentially expressed genes have no functional annotation: MGC4170 protein and hypothetical protein LOC283680 were down-regulated in all three samples while hypothetical protein BC012317 and hypothetical protein MGC39372 were up-regulated. Hypothetical protein MGC39372 showed an up-regulation at all three time-points in all three samples up to 8 fold at 24 hours and with a high significance ($p < 0.01$).

Next to the genes of so far unknown function three genes showed an up-regulation after 6 hours of induction. Trinucleotide repeat containing 9 or TOX (NP_001073899.1) which plays a role in the regulation of transcription by binding DNA, was up-regulated. Further a kinetochor associated protein 2 and a leukocyte immunoglobulin-like receptor subfamily B member 1. The direct effects of Rifampicin induction were not obviously detectable in the expression profile after 6 h.

24 h after adding Rifampicin to the cells the effect of induction was clearly represented in the expression profile. The volcano plot of this profile is shown in Figure 79.

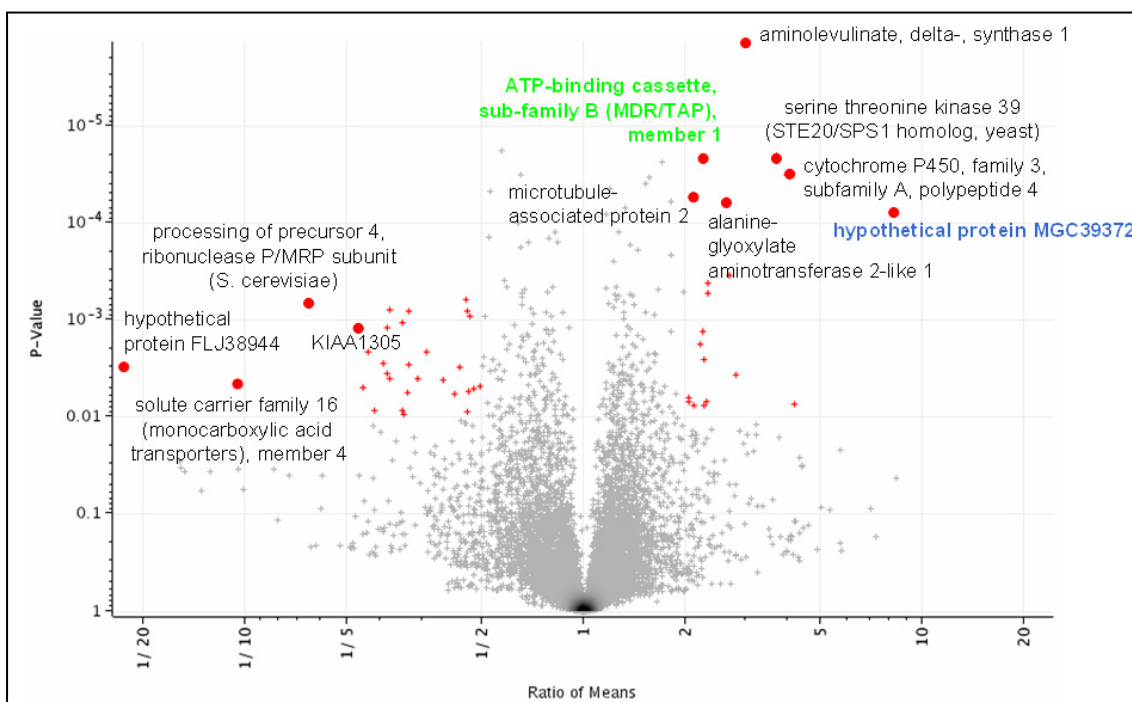


Figure 79: Volcano plot of expression data from samples treated 24 h with Rifampicin. 51 genes shown in red matching the criteria (2 fold and $p = 0.01$). Eleven genes with the most significant (low p-value) or highest change in expression level are given with their associated name/function.

Again several genes of unclear function including the previously up-regulated hypothetical protein MGC39372 were differentially expressed. After 24 h of induction also several genes involved in detoxification were induced:

The most significant up-regulated gene was the aminolevulinate delta-synthase I. The enzyme coded by this gene catalyzes the first step in the heme biosynthetic pathway. This observation was in good correlation with previous reports. The activity of this enzyme ensures that enough heme was supplied for the building of CYPs as they contain heme in their active centre. One of these CYPs, CYP3A4 was also highly significant up-regulated in all samples. Furthermore a transporter ATP-binding cassette sub-family B member 1 (ABCB1) was up-regulated (shown in green) and serine threonine kinase 39, microtubule-associated protein 2 and alanine-glyoxylate aminotransferase 2-like 1. On the other side a transporter solute carrier family 16 member 4 was down-regulated as well as precursor 4, ribonuclease P/MRP subunit, and KIAA1305.

The volcano plot of the expression values after 72 hours is given in Figure 80.

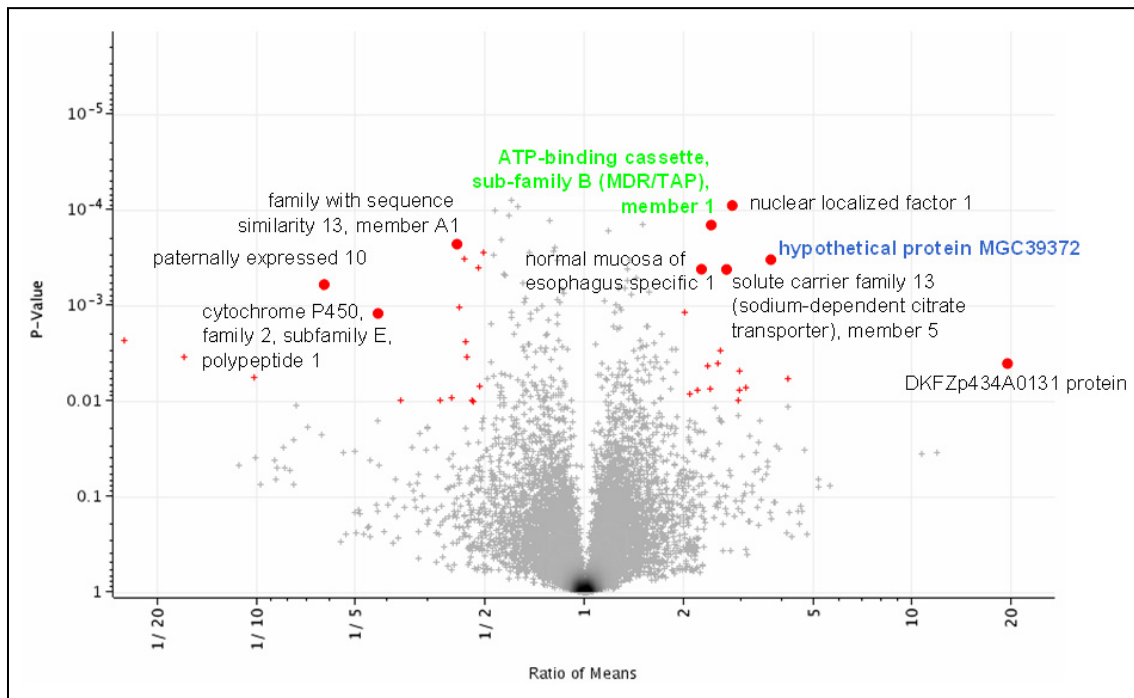


Figure 80: Volcano plot of expression data from samples treated 72 h with Rifampicin. 36 genes shown in red matching the criteria (2 fold and $p = 0.01$). Nine genes with the most significant (low p -value) or highest change in expression level are given with their associated name/function.

ATP-binding cassette sub-family B member 1 (ABCB1) was again significantly up-regulated after 72 hours of induction. As in the two previous time points the hypothetical protein MGC39372 was up-regulated in all three samples. Solute carrier family 13 member 5 shows a significant up-regulation while CYP2E1 was significantly down-regulated.

To structure the large amount of expression data several different clustering methods were applied.

3.8.1.2 Clustering

To group microarray data based on similar characteristics, clustering methods use distance functions to separate gene expression values into different categories. This grouping often allows assignment of functional roles and mapping to specific biological functions and processes. As the performance of the used tools, as well as the results, like the graphical representation depend on the selection of the distance function several

different functions were used. The aim was to identify co-regulation based on a positive correlation approach.

One kind of clustering is the so called principle component analysis (PCA). The general purpose of PCA is reduction of dimensionality. In microarray studies, the expression of a large number of genes is obtained for several samples. Each gene may be considered as a variable, or component of the individual cell culture. If only two genes and treatments were measured, a simple scatter plot of the values for each individual would reveal the influence of treatment. However the correlation to time and individual expression can not be visualized. For the analysis of larger numbers of genes, treatments and time points, graphical visualization of the resulting multidimensional space is impossible. Principal components analysis is a method for reducing the multidimensional space down to only three dimensions (details see 2.10.9).

The PCA-Plot shown in Figure 81 was performed for all genes and all arrays (three time points, two treatments).

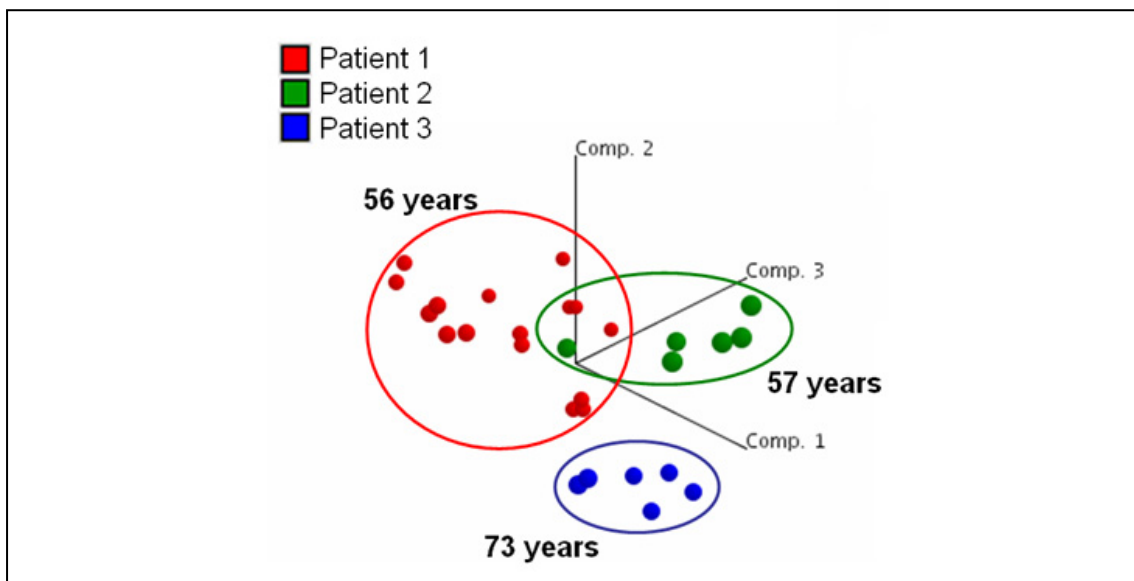


Figure 81: Principle component analysis (PCA) for all valid genes (approximately 17.000) and experiments (30). Shown are the different patients (colour coded) in a three dimensional space. Every ball indicates a microarray experiment from an individual RNA sample. The red balls indicating samples from patient 1 (HH26) as well as the green (patient 2, HH27) and blue (patient 3, HH44) form a cluster indicating that the individuality has a stronger impact on relationship than e.g. the treatment or time effects.

The spheres represent the expression profile of one array. It can be seen, if the expression of all genes is taken into consideration the three patients are clearly separated (Figure 81). Each patient shows a unique overall expression pattern. However if only for example genes involved in detoxification and xenobiotic metabolism (like CYPs) are used as input for the analysis, a separation between treated (Rifampicin) and not treated (control) samples can be found for all patients (Figure 82).

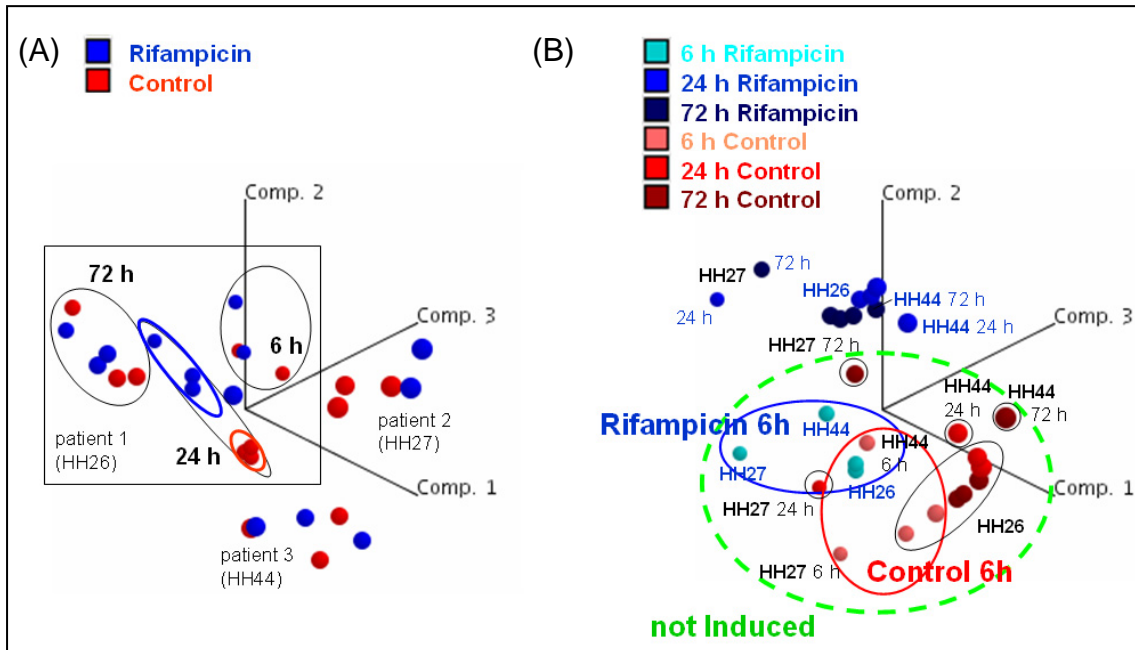


Figure 82: Principle component analysis (PCA) of all experiments for (A) all valid genes and (B) based on differentially expressed genes (CYP3A4 and ADH1 α). Control experiment without inducer are shown in red, arrays on which RNA from induced samples was hybridized are shown in blue. For the experiment with replicate arrays (HH26) indicated by the rectangle in part (A) it can be seen that after 24 hours there is a clear separation between the three control samples (red) which built a cluster and three test samples (blue). For the 6 h and the 72 h time point this separation is not as clear. In part (B) a similar plot was done based on CYP3A4 and ADH1a, however in this case the different time points were taken into account. The not induced samples (dashed green line) cluster apart from the induced (Rifampicin) ones (blue). In the not induced cluster the 6 h induced samples (blue) can be found, indicating that the induction process had no or at least a less dominant effect at this early point in time.

In part A of Figure 82 the control arrays are shown in red, while the arrays hybridized with Rifampicin treated samples are given in blue. As there were triplicate measurements performed for patient 1 (HH26) which cluster together, the separation between control and induced samples can be seen very clearly. Especially after 24 h there is a strong difference of the expression profile between induced (blue circle) and control samples (red circle).

The time dependent expression profile was analysed in more detail (part B of Figure 82). The PCA plot revealed a cluster containing the not induced samples. This cluster is marked by a green dashed line. In this control-cluster three Rifampicin induced (blue) samples were found. These three arrays represent the expression level of the 6 h samples from patient 1, 2 and 3. As they are clustered to the not induced samples there expression profile was more correlated to not induced samples. This analysis showed that for the investigation of genes differentially expressed by Rifampicin the 24 hours time point and the 72 h time point were more important than the 6 h time point. After 6 h there was no strong expression change between Rifampicin and control samples. For subsequent analysis, especially for network reconstruction the focus was put on the differentially expressed genes after 24 h and 72 h of induction.

To identify gene clusters which were affected by administering Rifampicin a 2D-clustering was performed. The results of this clustering are shown in Figure 83. In part A of Figure 83 an overview of the colour coded expression values is given. Red rectangles shows highly expressed genes and green low expressed genes. Each rectangle represents the expression value of one gene on one array. The arrays are clustered in the second dimension (on top). The array/experiment cluster is shown in part (B) of Figure 83. Three functional gene clusters showing a strong change in expression between Rifampicin treated and control cells are also given enlarged in Figure 83 section (C), (D) and (E).

Because PCA had revealed the strongest effect of Rifampicin 24 h after induction especially the expression of genes at this measurement point was investigated. The experiments clustered mainly based on their treatment (Figure 83 B). One Rifampicin and one control cluster were obtained for the arrays analysed here containing samples 24 h after induction. Sample HH27 showed a different behaviour. The control and induced samples from this patient clustered separately together. This cluster showed a higher correlation to the induced samples compared to the control samples.

In Figure 83 C the expression profile of five metallothionein 1 (*mt1*) genes is shown. MT1F, K, X, H and E are high expressed in all three patients. The expression in the Rifampicin treated cells was higher compared to DMSO control samples.

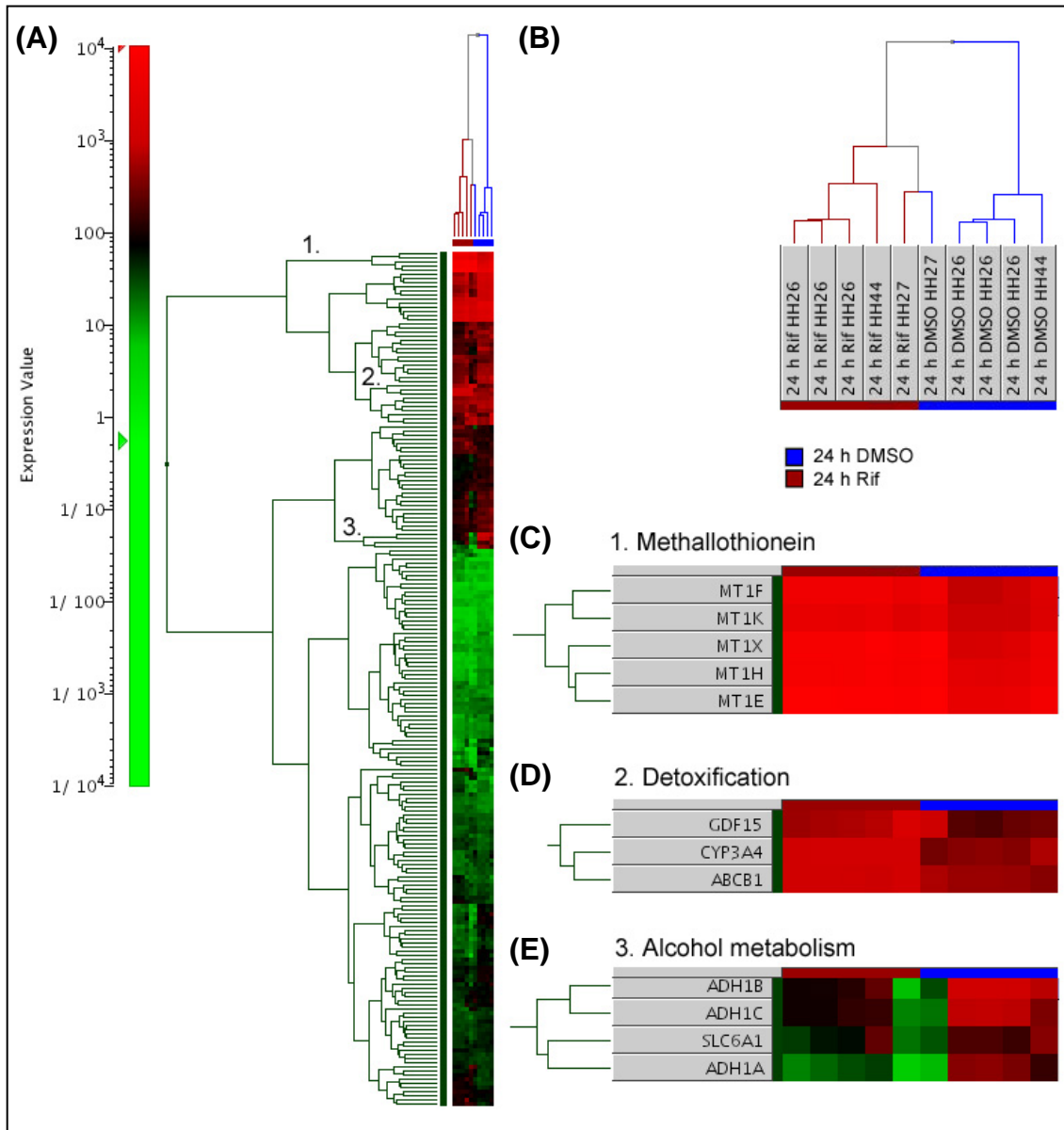


Figure 83: 2D-hierarchical cluster of experiments and expressed genes in human hepatocytes after 24 h induction with Rifampicin. The expression level is colour coded: green represents low expression and red high expression. (A) presents an overview of all clustered differentially expressed genes. The levels of expression are given for all five DMSO (blue) and five Rifampicin (red) arrays. In (B) the experimental clustering is shown. The three arrays hybridized with RNAs from cell culture HH26 cluster together, representing their close relationship (technical replicates). Overall the rifampicin arrays build one cluster in contrast to the second DMSO cluster. The HH27 DMSO array however clusters to the Rifampicin samples. Three gene clusters were shown enlarged (C - E). (C) Shows the highly expressed methallothionein genes forming one cluster. (D) The second "detoxification" cluster consists of CYP3A4 and ABCB1 as well as GDF15. In third cluster (E) members of the alcohol metabolism were found: ADH18, AH1C and ADH1A. Furthermore SLC6A1 fitted well in the "alcohol metabolism" cluster.

The second up-regulated cluster shown in Figure 83 D was named "detoxification" as cytochrome P450, family 3, subfamily A, polypeptide 4 (CYP3A4) and ATP-binding cassette sub-family B member 1 (ABCB1) clustered together in this group. In addition growth differentiation factor 15 (GDF15) was similarly expressed with the two other genes directly involved in detoxification. In this cluster samples from cell culture HH27 showed the same expression in control samples as in induced ones.

Shown in part E of Figure 83 a cluster with genes involved in alcohol metabolism is given. This cluster consists of alcohol dehydrogenase (class I), alpha polypeptide (ADH1A), alcohol dehydrogenase (class I), beta polypeptide (ADH1B), alcohol dehydrogenase 1C (class 1), gamma polypeptide (ADH1C) and additionally solute carrier family 6 (neurotransmitter transporter, GABA), member 1 (SLC6A1). This cluster was highly expressed in the control cells and low expressed in the Rifampicin cells. In samples from patient HH27 the genes were low expressed in both, Rifampicin as well as control DMSO cells.

Using tree and tile plot visualization and item clustering with a distance maximum, complete measurement and a valid value score of 50% the following cluster diagrams for the 24 h samples were obtained (Figure 84 - Figure 87).

Up-regulated cluster 1 consisted of alanine-glyoxylate aminotransferase 2-like 1 (AGXT2L1), CD14 antigen (CD14) which is a surface protein preferentially expressed on monocytes/macrophages. This antigen binds the lipopolysaccharide binding protein and has been shown to bind apoptotic cells. 5-aminolevulinate synthase (ALAS1) also in this cluster is known to be involved in the porphyrin metabolism. A high correlation to this group, concerning the expression profile, was observed for the down-regulated gene serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10 (SERPINA10).

In the second up-regulated cluster three metallothionein genes (MT1X, MT1K and MT1F), which might be involved in regulation of cell proliferation and apoptosis and in various other physiological processes and cytochrome P450, family 2, subfamily C, polypeptide 8 (CYP2C8) were grouped.

The third cluster, which was also up-regulated, included several genes involved in detoxification. Cytochrome P450, family 3, subfamily A, polypeptide 4 (CYP3A4),

cytochrome P450, family 2, subfamily C, polypeptide 19 (CYP2C19), cytochrome P450, family 2, subfamily A, polypeptide 6 (CYP2A6), ATP-binding cassette subfamily B member 1 and growth differentiation factor 15 (GDF15) were summarized in the third cluster. GDF15 belongs to the group of morphogenetic proteins which are members of the transforming growth factor-beta superfamily and regulate tissue differentiation and maintenance.

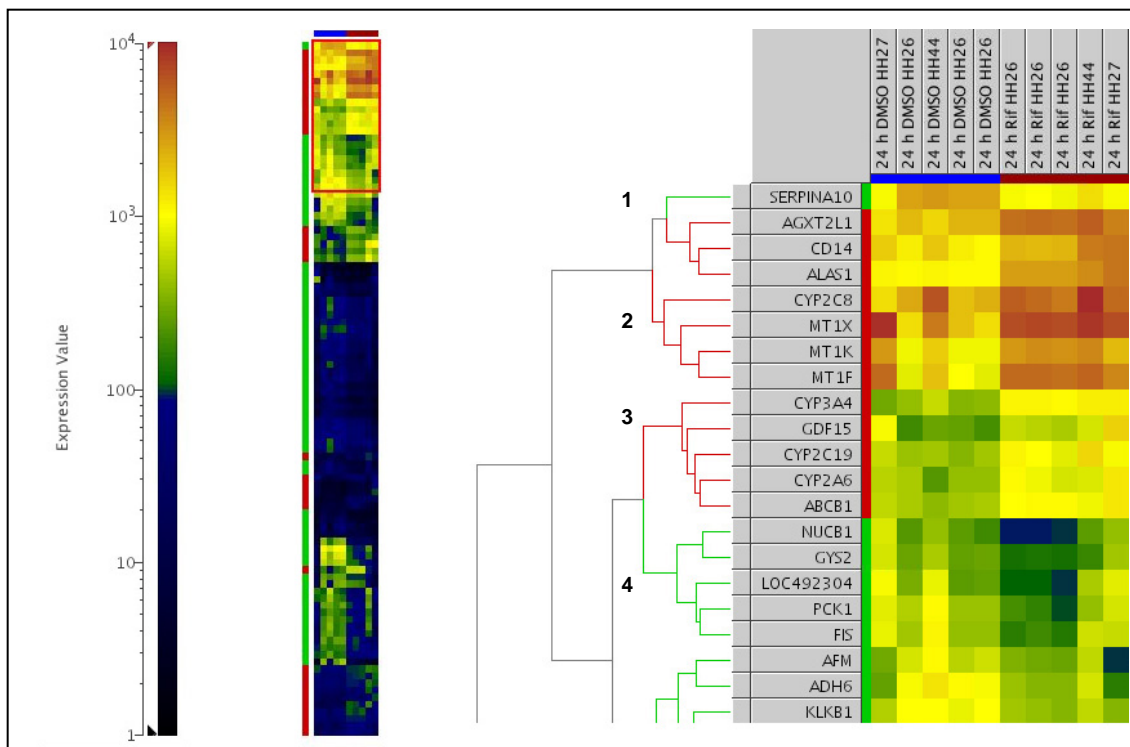


Figure 84: Global gene expression profiles for the 2-fold differentially expressed genes in the 24 h samples of control (blue) and DMSO (red) samples. The branches of the dendrogram generated by maximum, complete, 50 % item clustering reveal coherence in the expression data. Green lines connect down-regulated genes after inducing with Rifampicin. The red lines indicate up-regulation after this stimulus. The colour coded tile plot gives the expression strength of the according gene (yellow red stands for high expression, green blue for low expression). Enlarged is the upper part of the global expression clustering profile indicated by the red square. Three up-regulated clusters (1, 2 and 3) and one down-regulated (4) cluster are shown here.

The fourth cluster combining similarly 2-fold down-regulated genes consisted of nucleobinding 1 (NUCB1, CG32190-PA), glycogen synthase 2 from the liver (GYS2), a putative insulin-like growth factor II associated protein (LOC492304), phosphoenolpyruvate carboxykinase 1 (PCK1) and chromosome 5 open reading frame 27 (FIS) of unknown function. PCK1, the main control point for the regulation of

gluconeogenesis was down-regulated. The cytosolic enzyme encoded by this gene, along with GTP, catalyzes the formation of phosphoenolpyruvate from oxaloacetate, with the release of carbon dioxide and GDP. The expression of this gene can be regulated by insulin, glucocorticoids, glucagon, cAMP, and diet. Furthermore GYS2 was found in the same cluster so the Insulin signalling pathway and metabolism of carbohydrates seems to be affected by stimulating cells with Rifampicin.

In the second part of the global gene expression profile (Figure 85) one up-regulated cluster (6) and two down-regulated clusters (5 and 7) were obtained. The following genes cluster together in the group of up-regulated gene expression: trinucleotide repeat containing 9 (TNRC9), cytochrome P450, family 2, subfamily A, polypeptide 7 (CYP2A7), cytochrome P450, family 2, subfamily B, polypeptide 6 (CYP2B6), cytochrome P450, family 3, subfamily A, polypeptide 43 (CYP3A43) and proline dehydrogenase (oxidase) 2 (PRODH2). The down-regulated cluster 1 comprises the following genes: afamin (AFM) which is a member of the albumin gene family. The protein encoded by this gene is regulated developmentally, expressed in the liver and secreted into the bloodstream. Further genes in this cluster are: alcohol dehydrogenase 6 class V (ADH6), kallikrein B, plasma (Fletcher factor) 1 (KLKB1), hydroxyacid oxidase 2 - long chain (HAO2), acyl-coA synthetase medium-chain family member 5 (ACSM5, FLJ20581), alcohol dehydrogenase 4 (class II), pi polypeptide (ADH4), cytochrome P450, family 2, subfamily E, polypeptide 1 (CYP2E1) and zinc finger, CCCH-type with G patch domain (ZGPAT).

adh6 gene encodes class V alcohol dehydrogenase, which is a member of the alcohol dehydrogenase family. Members of this family metabolize a wide variety of substrates, including ethanol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products. This gene is expressed in the stomach as well as in the liver, and it contains a glucocorticoid response element upstream of its 5' UTR, which is a steroid hormone receptor binding site. KLKB1 is a glycoprotein that participates in the surface-dependent activation of blood coagulation, fibrinolysis, kinin generation and inflammation. It is synthesized in the liver and secreted into the blood as a single polypeptide chain.

hao2 is one of three related genes that have 2-hydroxyacid oxidase activity. The sub-cellular location of the encoded protein is the peroxisome. Specifically, this gene is

expressed predominantly in liver and kidney and has the highest activity toward the substrate 2-hydroxypalmitate. ACSM5 is as acyl-CoA synthetase involved in lipid metabolism and secondary metabolites biosynthesis and catabolism.

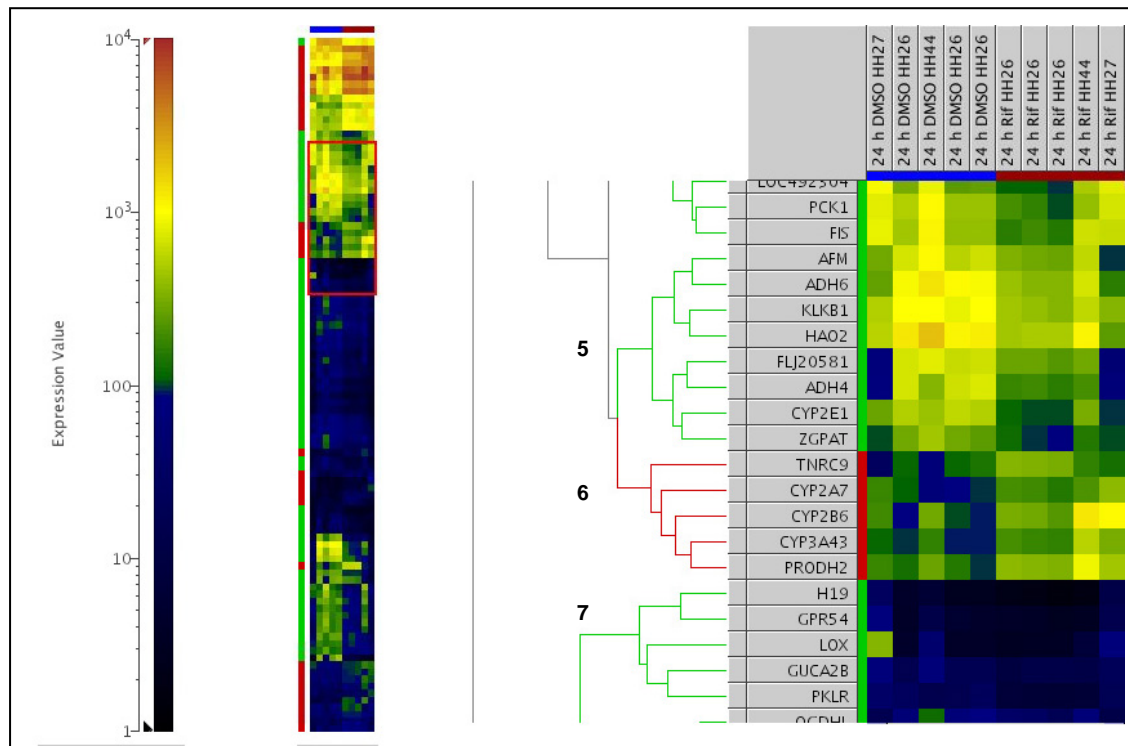


Figure 85: Global gene expression profiles for the 2-fold differentially expressed genes in the 24 h samples of control (blue) and DMSO (red) samples (part 2). The branches of the dendrogram generated by maximum, complete, 50 % item clustering reveal coherence in the expression data. Green lines connect down-regulated genes after inducing with Rifampicin. The red lines indicate up-regulation after this stimulus. The colour coded tile plot gives the expression strength of the according gene (yellow red stands for high expression, green blue for low expression). Enlarged is the upper part of the global expression clustering profile indicated by the red square. One up-regulated cluster (6) and two down-regulated (5 and 7) clusters are shown here.

adh4 gene encodes class II alcohol dehydrogenase 4 subunit, which is a member of the alcohol dehydrogenase family. Members of this enzyme family metabolize a wide variety of substrates, including ethanol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products. Class II alcohol dehydrogenase is a homodimer composed of 2 subunits. It exhibits a high activity for oxidation of long-chain aliphatic alcohols and aromatic alcohols.

CYP2E1 protein is a member of the cytochrome P450 monooxygenase family which catalyzes many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. This protein localized to the endoplasmic reticulum and is

induced by ethanol, the diabetic state, and starvation. The enzyme metabolizes both endogenous substrates, such as ethanol, acetone, and acetal, as well as exogenous substrates including benzene, carbon tetrachloride, ethylene glycol, and nitrosamines which are pre-mutagens found in cigarette smoke. Due to its many substrates, this enzyme may be involved in such varied processes as gluconeogenesis, hepatic cirrhosis, diabetes, and cancer.

The second down-regulated cluster consisted of five poorly expressed genes: H19 fetal liver mRNA (H19, imprinted maternally expressed untranslated mRNA), G protein-coupled receptor 54 equal to KISS1 receptor (GPR54, KISS1R), lysyl oxidase (LOX), guanylate cyclase activator 2B (GUCA2B) and pyruvate kinase, liver (PKLR).

The protein KISS1 receptor is a galanin-like G protein-coupled receptor that binds metastin, a peptide encoded by the metastasis suppressor gene KISS1. The tissue distribution of the expressed gene suggests that it is involved in the regulation of endocrine function. The protein encoded by the gene *lox* is an extracellular copper enzyme that initiates the cross-linking of collagens and elastin. The enzyme catalyzes oxidative deamination of the epsilon-amino group in certain lysine and hydroxylysine residues of collagens and lysine residues of elastin. In addition to cross-linking extracellular matrix proteins, the encoded protein may have a role in tumour suppression. Guanylate cyclase activator 2B (GUCA2B) is a peptide homolog of the bacterial heat-stable enterotoxins (e.g., the *E. coli* ST toxin). It is an endogenous activator of the guanylate cyclase-2C receptor, which synthesizes cyclic GMP (cGMP), a key component of several intracellular signal transduction pathways. The protein encoded by the gene *pklr* is a pyruvate kinase that catalyzes the production of phosphoenolpyruvate from pyruvate and ATP.

The first cluster (8) of Figure 86 includes the down-regulated nuclear receptor NR1I3 (CAR) which is a key regulator of xenobiotic and endobiotic metabolism. The protein binds to DNA as a monomer or a heterodimer with the retinoid X receptor (RXR) and regulates the transcription of target genes involved in drug metabolism and bilirubin clearance, such as cytochrome P450 family members. Unlike most nuclear receptors, this transcriptional regulator is constitutively active in the absence of ligand but is regulated by both agonists and inverse agonists. Ligand binding results in translocation of this protein to the nucleus, where it activates or represses target gene transcription.

These ligands include bilirubin, a variety of foreign compounds, steroid hormones, and prescription drugs. NR1I3 forms the first cluster (8) with the so far unknown and not annotated LOC255167. In the next higher branch LOC343071 is clustered to this group.

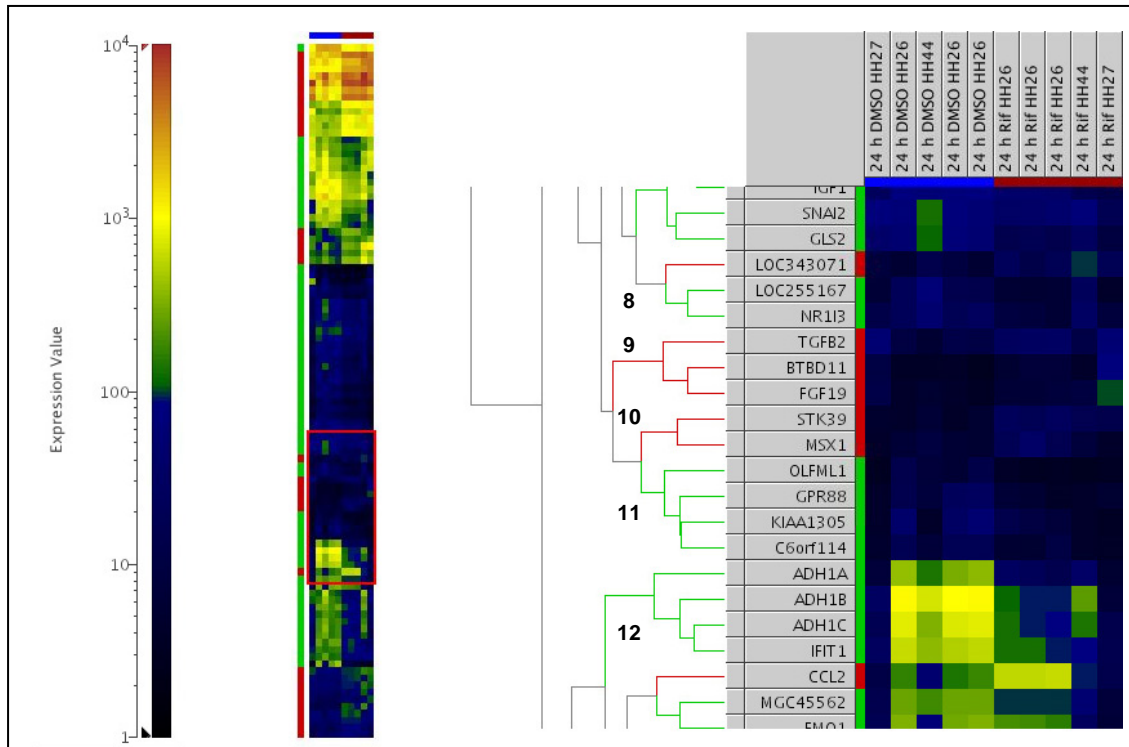


Figure 86: Global gene expression profiles for the 2-fold differentially expressed genes in the 24 h samples of control (blue) and DMSO (red) samples (part 3). The branches of the dendrogram generated by maximum, complete, 50 % item clustering reveal coherence in the expression data. Green lines connect down-regulated genes after inducing with Rifampicin. The red lines indicate up-regulation after this stimulus. The colour coded tile plot gives the expression strength of the according gene (yellow red stands for high expression, green blue for low expression). Enlarged is the lower part of the global expression clustering profile indicated by the red square. Two up-regulated cluster (9 and 10) and three down-regulated clusters (8, 11 and 12) are shown here.

Genes from the second cluster were up-regulated in Rifampicin treated cells compared to control cells. Three genes form this cluster: transforming growth factor, beta 2 (TGFB2), BTB (POZ) domain containing 11 (BTBD11) and fibroblast growth factor 19 (FGF19).

TGFB2 is a secreted protein known as a cytokine that performs many cellular functions and has a vital role during embryonic development. It is known to suppress the effects of interleukin dependent T-cell tumours. FGF19 is a member of the fibroblast growth factor (FGF) family. Members possess broad mitogenic and cell survival activities, and

are involved in a variety of biological processes including embryonic development cell growth, morphogenesis, tissue repair, tumour growth and invasion. The third cluster is formed by two genes showing a similar expression: serine threonine kinase 39 (STK39) and msh homeobox 1 (MSX1). The role of STK39 was described previously. MSX1 protein functions as a transcriptional repressor during embryogenesis through interactions with components of the core transcription complex and other homeoproteins. It may also have roles in limb-pattern formation, craniofacial development, particularly odontogenesis, and tumour growth inhibition.

Cluster 4 containing the following genes was down-regulated: olfactomedin-like 1 (*olfml1*), G protein-coupled receptor 88 (*gpr88*) and *kiaa1305* chromosome 6 open reading frame 114 (*c6orf114*). Genes in this cluster showed a low expression for both treatments. In the fifth cluster genes grouped together have a high expression in the DMSO treated samples (except for culture from patient HH27) and are low expressed in the samples treated with Rifampicin. This cluster summarizes alcohol dehydrogenase 1A (class I), alpha-, beta- and gamma polypeptide (ADH1A, B, C) and interferon-induced protein with tetratricopeptide repeats (1IFIT1).

Two clusters are formed in the lower part of the global clustering and displayed in Figure 87. Cluster number 13 again combines genes which are highly expressed in DMSO control cells and down-regulated in Rifampicin treated cells. This down-regulation was also observed for cell culture from patient 44. However this patient had an overall higher expression of the clustered genes.

Cytochrome P450, family 1, subfamily A, polypeptide 2 (CYP1A2), retinol dehydrogenase 16 (RODH-4), hypothetical protein LOC149703 (LOC149703), ASCL830 (UNQ830), amino-methyl-transferase (AMT), solute carrier family 6 (neurotransmitter transporter, GABA), member 1 (SLC6A1), secreted phosphoprotein 2 (SPP2, SPP24), collagen, type V, alpha 2 (COL5A2), chromosome 14 open reading frame 94 (C14orf94) and mannose receptor, C type 1 (MRC1) belong to this down-regulated cluster. CYP1A2 protein localizes to the endoplasmic reticulum and its expression is induced by some polycyclic aromatic hydrocarbons (PAHs), some of which are found in cigarette smoke. The enzyme's endogenous substrate is unknown. However CYP1A2 is able to metabolize some PAHs to carcinogenic intermediates.

Other xenobiotic substrates for this enzyme include caffeine, aflatoxin B1, and acetaminophen.

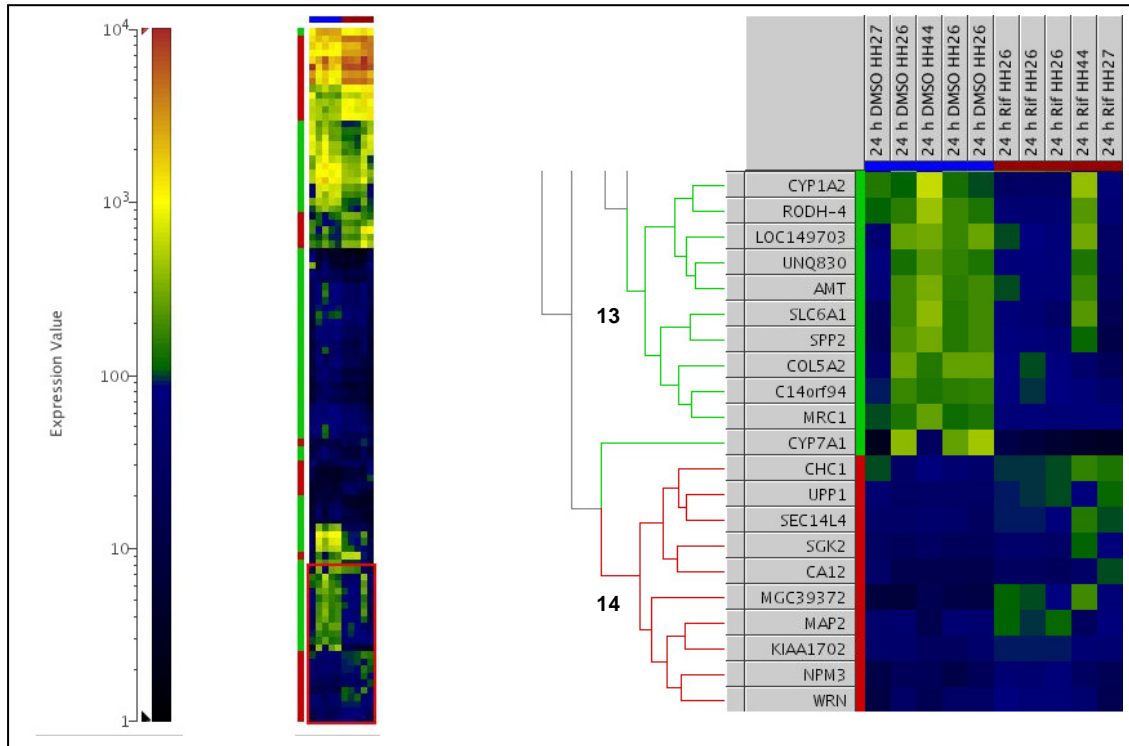


Figure 87: Global gene expression profiles for the 2-fold differentially expressed genes in the 24 h samples of control (blue) and DMSO (red) samples (part 4). The branches of the dendrogram generated by maximum, complete, 50 % item clustering reveal coherence in the expression data. Green lines connect down-regulated genes after inducing with Rifampicin. The red lines indicate up-regulation after this stimulus. The colour coded tile plot gives the expression strength of the according gene (yellow red stands for high expression, green blue for low expression). Enlarged is the lower part of the global expression clustering profile indicated by the red square. One up-regulated cluster (14) and one down-regulated cluster (13) are shown here.

AMT is an enzyme which catalyzes the creation of methylene-tetrahydrofolate. It is part of the glycine decarboxylase complex. COL5A2 encodes an alpha chain for one of the low abundance fibrillar collagens. Fibrillar collagen molecules are trimers that can be composed of one or more types of alpha chains. Type V collagen is found in tissues containing type I collagen and appears to regulate the assembly of heterotypic fibers composed of both type I and type V collagen. MRC1 protein is a type I membrane receptor that mediates the endocytosis of glycoproteins by macrophages. The protein has been shown to bind high-mannose structures on the surface of potentially pathogenic viruses, bacteria, and fungi so that they can be neutralized.

The second cluster (14) showed an overall low expression. Several genes showed a bit higher expression in Rifampicin treated samples compared to controls. This cluster (14) consists of the following slightly up-regulated genes: regulator of chromosome condensation 1 (*chc1*, *rcc1*), uridine phosphorylase 1 (*upp1*), SEC14-like 4 - *S. cerevisiae* (*sec14l4*), serum/glucocorticoid regulated kinase 2 (*sgk2*), carbonic anhydrase XII (*ca12*), hypothetical protein MGC39372 (*mgc39372*), microtubule-associated protein 2 (*map2*), KIAA1702 protein (*kiaa1702*), nucleophosmin/nucleoplasmin, 3 (*npm3*) and Werner syndrome (*wrn*).

SEC14L4 is highly similar to the protein encoded by the *Saccharomyces cerevisiae* SEC14 gene. The SEC14 protein is a phosphatidylinositol transfer protein that is essential for biogenesis of Golgi-derived transport vesicles, and thus is required for the export of yeast secretory proteins from the Golgi complex. The specific function of this protein in humans has not yet been determined. *sgk2* gene encodes a serine/threonine protein kinase. Although this gene product is similar to serum- and glucocorticoid-induced protein kinase (SGK), this gene is not induced by serum or glucocorticoids. It is induced in response to signals that activate phosphatidylinositol 3-kinase, which is also true for SGK. Carbonic anhydrases (CAs, e.g. CA12) are a large family of zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide. They participate in a variety of biological processes, including respiration, calcification, acid-base balance, bone resorption, and the formation of aqueous humor, cerebrospinal fluid, saliva, and gastric acid. This gene product is a type I membrane protein that is highly expressed in normal tissues, such as kidney, colon and pancreas, and has been found to be over-expressed in 10 % of clear cell renal carcinomas. MAP2 protein belongs to the microtubule-associated protein family. The proteins of this family are thought to be involved in microtubule assembly, which is an essential step in neurogenesis. The products of similar genes in rat and mouse are neuron-specific cytoskeletal proteins that are enriched in dendrites, implicating a role in determining and stabilizing dendritic shape during neuron development. NPM3 is related to the nuclear chaperone phosphoproteins, nucleoplasmin and nucleophosmin. It is highly homologous to the murine *Npm3* gene. Based on the structural similarity of the human NPM3 gene product to nucleoplasmin and nucleophosmin, NPM3 may represent a new member of this gene family, and may share basic functions with the molecular chaperones. WRN is

a member of the RecQ subfamily and the DEAH (Asp-Glu-Ala-His) subfamily of DNA and RNA helicases. DNA helicases are involved in many aspects of DNA metabolism, including transcription, replication, recombination, and repair. This protein possesses an intrinsic 3' to 5' DNA helicase activity, and is also a 3' to 5' exonuclease. Based on interactions between this protein and Ku70/80 heterodimer in DNA end processing, this protein may be involved in the repair of double strand DNA breaks. Defects in this gene are the cause of Werner syndrome, an autosomal recessive disorder characterized by premature aging. This gene as the other ones mentioned previously from this cluster are all low expressed in the human hepatocytes after 24 hours. Therefore their role in detoxification should not be overestimated.

3.8.2 Gene Expression Profiling Based on Cellular Function

On one hand genes can be grouped by similar expression. This was used by the clustering method in the previous part. On the other hand expression profiles for genes with known similar cellular function can be grouped. This allows investigation of expression changes in groups sharing common cellular functions. Ratios of genes differentially expressed in Rifampicin and control cells were calculated. Groups with similar functions like apoptosis, cell adhesion, cell cycle control, cell signalling, drug metabolism and transport, lipid metabolism and housekeeping are plotted in Figure 88 - Figure 93.

In Figure 88 the expression of several genes involved in apoptosis and cell adhesion are given: BCL2-antagonist of cell death (BAD) was up-regulated after 24 and 72 h. *bad* is a member of the BCL-2 family. BCL-2 family members are known to be regulators of programmed cell death. This protein positively regulates cell apoptosis by forming heterodimers with BCL-xL and BCL-2, and reversing their death repressor activity. Proapoptotic activity of this protein is regulated through its phosphorylation. Protein kinases AKT and MAP kinase, as well as protein phosphatase calcineurin were found to be involved in the regulation of this protein. BCL-2-associated X protein (BAX) also shown, forms a heterodimer with BCL2, and functions as an apoptotic activator. This protein is reported to interact with, and increase the opening of, the mitochondrial voltage-dependent anion channel (VDAC), which leads to the loss in membrane

potential and the release of cytochrome c. The expression of this gene is regulated by the tumour suppressor P53 and has been shown to be involved in P53-mediated apoptosis. After 24 h of Rifampicin induction the gene coding BAX showed an almost 1.5-fold increase in expression. After 72 h the expression declined to 1.2-fold. Caspase 7 and 8, apoptosis-related cysteine peptidases (CASP7 and 8) are members of the cysteine-aspartic acid protease (caspase) family. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis. Caspases exist as inactive pro-enzymes which undergo proteolytic processing at conserved aspartic residues to produce two subunits that dimerize to form the active enzyme. It is activated upon cell death stimuli and induces apoptosis. In this experiment the caspases showed no differentially expression over time between Rifampicin and control samples.

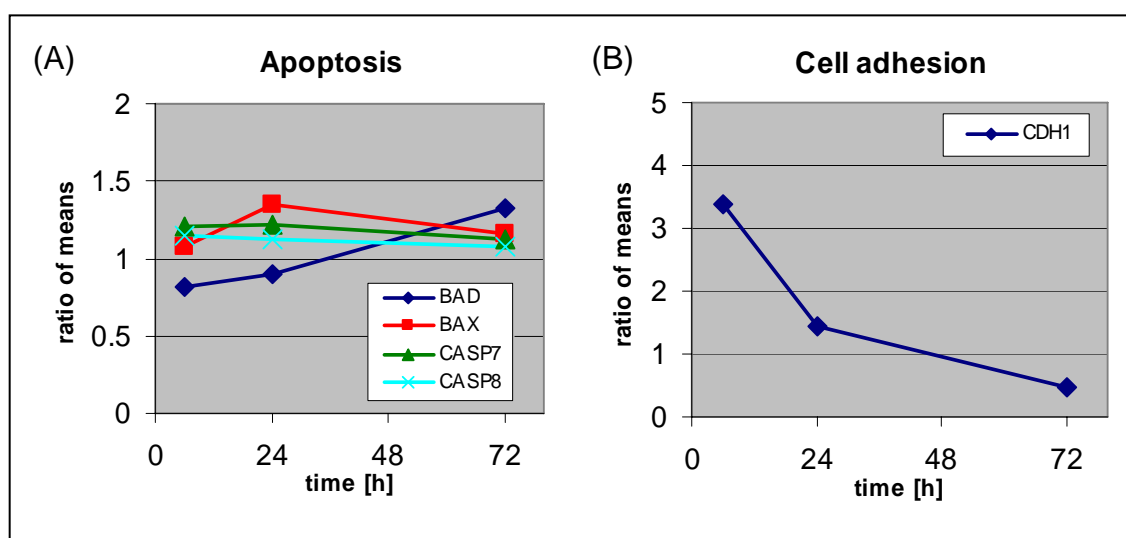


Figure 88: Expression profiles of genes involved in (A) apoptosis and (B) cell adhesion. In part (A) the expression ratios for BAD, BAX, CASP7 and CASP8 are plotted over time. In part (B) the expression ratios between DMSO control samples and Rifampicin treated cells after 6, 24 and 72 hours are given.

Cadherin 1, type 1, E-cadherin (CDH1) showed a remarkable down-regulation over time from a 3.5-fold over-expression after 6 h of induction to 1.5-fold over-expression after 24 h to a 2-fold down-regulation after 72 h of induction. The encoded protein is a calcium dependent cell-cell adhesion glycoprotein comprised of five extracellular cadherin repeats, a transmembrane region and a highly conserved cytoplasmic tail. Loss of function is thought to contribute to progression in cancer by increasing proliferation, invasion, and/or metastasis.

In Figure 89 genes involved in cell cycle regulation are shown. Cyclin A2 (CCNA2) and cyclin B1 (CCNB1) showed a similar expression profile. CCNA2 had a higher increase after 24 h of Rifampicin induction compared to CCNB1 which was also induced. After 72 h both genes were declining to the ratio they had after 6 h (CCNA2) or even lower (CCNB1).

CCNA2 belongs to the highly conserved cyclin family, whose members are characterized by a dramatic periodicity in protein abundance through the cell cycle. Cyclins function as regulators of CDK kinases. Different cyclins exhibit distinct expression and degradation patterns which contribute to the temporal coordination of each mitotic event. In contrast to cyclin A1, which is present only in germ cells, this cyclin is expressed in all tissues tested. This cyclin binds and activates CDC2 or CDK2 kinases, and thus promotes both cell cycle G1/S and G2/M transitions. CCNB1 is a regulatory protein involved in mitosis. Cyclin-dependent kinase 4 (CDK4) showed a slight induction after 24 and 72 hours of induction. CDK4 is a member of the Ser/Thr protein kinase family. It is a catalytic subunit of the protein kinase complex that is important for cell cycle G1 phase progression. All other genes in the graph "Cell cycle 1" (*ccnd1*, *ccnd3*, *cdk2*, *cdk6*, *cdkn1a*, *cdkn1b*) showed no differential expression or only very slight down-regulation.

Mitogen-activated protein kinase 1 (MAP2K1) is a member of the dual specificity protein kinase family, which acts as a mitogen-activated protein (MAP) kinase kinase. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act as an integration point for multiple biochemical signals. This protein kinase lies upstream of MAP kinases and stimulates the enzymatic activity of MAP kinases upon wide variety of extra- and intracellular signals. As an essential component of MAP kinase signal transduction pathway, this kinase is involved in many cellular processes such as proliferation, differentiation, transcription regulation and development. MAP2K1 was down-regulated 2-fold after 24 h of induction. After 72 h of induction the expression ratio turned back to the starting ratio after 6 h of induction. Comparable expression behaviour to MAP2K1 was observed for tumour protein p53 (TP53).

Tumour protein p53, a nuclear protein, plays an essential role in the regulation of cell cycle, specifically in the transition from G0 to G1 phase of the cycle. It is found in very low levels in normal cells. However in a variety of transformed cell lines, it is

expressed in high amounts, and therefore believed to contribute to transformation and malignancy. p53 is a DNA-binding protein containing DNA-binding, oligomerization and transcription activation domains. It is postulated to bind as a tetramer to a p53-binding site and activate expression of downstream genes that inhibit growth and/or invasion, and thus function as a tumour suppressor. Mutants of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA binding site, and hence cause the loss of tumour suppressor activity.

Telomerase-associated protein 1 (TEP1) which is a component of the ribonucleo-protein complex responsible for telomerase activity which catalyzes the addition of new telomeres on the chromosome ends was slightly down-regulated from 1.5-fold induction after 6 hours to 1.2-fold after 24 and 72 hours.

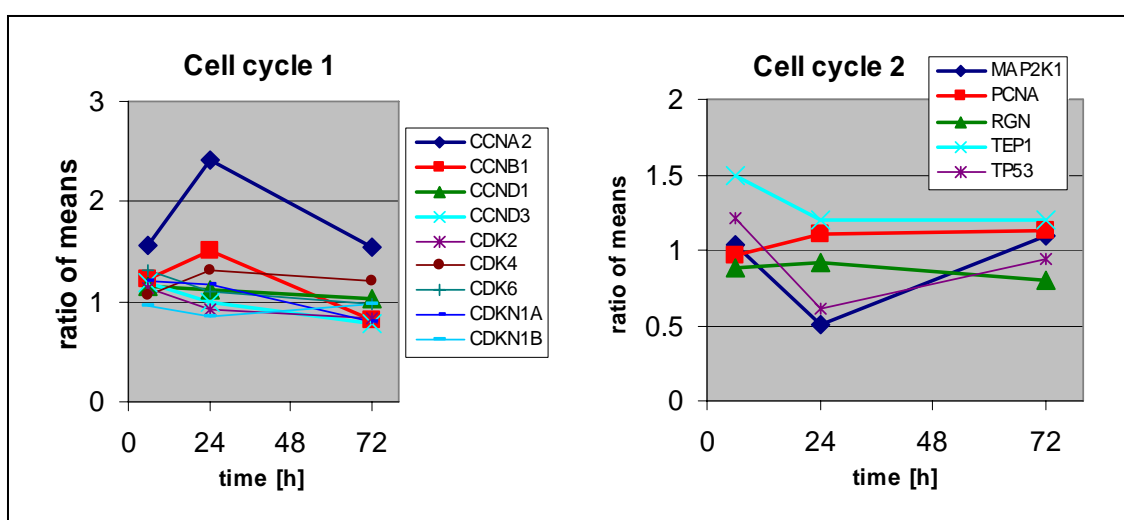


Figure 89: Expression profiles of genes involved in cell cycle regulation. The expression ratios between DMSO (control) and Rifampicin treated cells after 6, 24 and 72 hours are given for several cyclins, cyclin-dependent kinases and MAP2K1, PCNA, RGN, TEP1 and TP53.

Proliferating cell nuclear antigen (PCNA), which is a co-factor of DNA polymerase delta helping in the process of leading strand synthesis during DNA replication, as well as regucalcin (senescence marker protein-30; RGN) showed almost no change in their expression ratio. RGN is a highly conserved, calcium-binding protein that is preferentially expressed in the liver and kidney. It may have an important role in calcium homeostasis. Studies with rats indicate that this protein may also play a role in aging, as it shows age-associated down-regulation.

In concern of cell signalling IGF2R, IL6R and NFKB1 α were investigated. Insulin-like growth factor 2 receptor (IGF2R) functions in the intracellular trafficking of lysosomal enzymes, the activation of transforming growth factor beta, and the degradation of IGF2. After induction with Rifampicin no differential expression was observed in any point of time. Interleukin 6 receptor (IL6R) and nuclear factor kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (NFKB1 α) were down-regulated from about 1.2-fold over-expression after 6 h to 1.4-fold down-regulation after 24 h and almost no difference in expression for NFKB1 α and 2-fold down-regulation for IL6R after 72 h. Interleukin 6 (IL6) is a potent pleiotropic cytokine that regulates cell growth and differentiation and plays an important role in immune response. The protein encoded by the gene IL6R is a subunit of the receptor complex for IL6. The IL6 receptor is a protein complex consisting of this protein and interleukin 6 signal transducer (IL6ST/GP130/ IL6-beta), a receptor subunit also shared by many other cytokines. Disregulated production of IL6 and this receptor are implicated in the pathogenesis of many diseases, such as multiple myeloma, autoimmune diseases and prostate cancer.

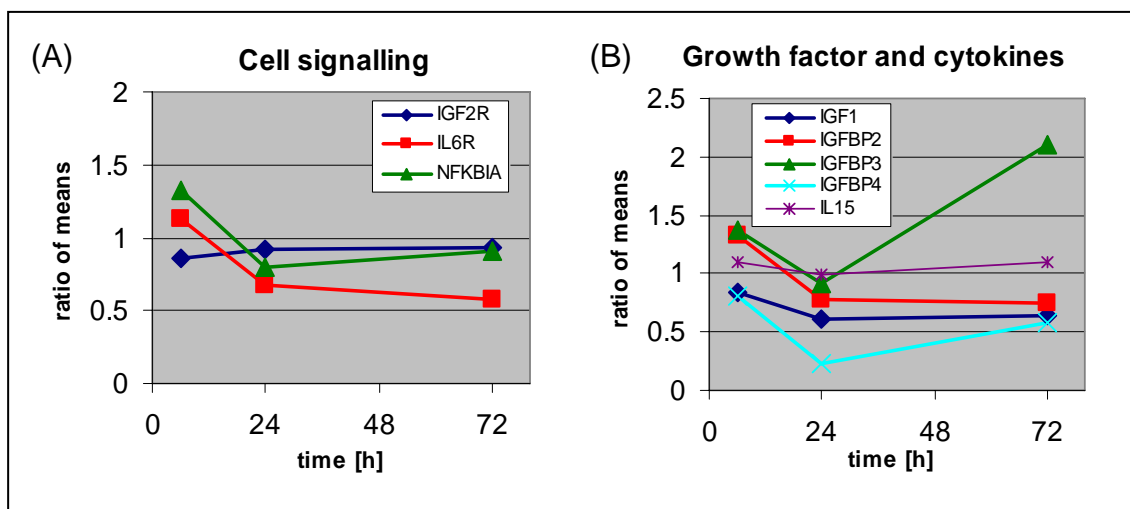


Figure 90: Expression profiles of genes involved in (A) cell signalling and (B) coding for growth factors and cytokines. In part (A) the expression ratios for IGF2R, IL6R and NFKB1A are plotted over time. In part (B) the expression ratios between DMSO control samples and Rifampicin treated cells after 6, 24 and 72 hours are given for IGF1, IGFBP2, IGFBP3, IGFBP4 and IL15.

Insulin-like growth factor 1 (IGF1) is one member of insulin-like growth factors (IGFs), which comprise a family of peptides that play important roles in mammalian growth and development. IGF1 mediates many of the growth-promoting effects of growth

hormone (GH). In this investigation it showed a 1.6-fold down-regulation after 24 and 72 hours. Insulin-like growth factor binding protein 2, 3 and 4 (IGFBP 2, 3 and 4) showed a variable down-regulation after 24 h of induction. While IGFBP 2 and 3 kept their expression ratio after 72 hours, IGFBP 3 showed a strong up-regulation of 2-fold. IGFBP3 is a member of the insulin-like growth factor binding protein (IGFBP) family and encodes a protein with an IGFBP domain and a thyroglobulin type-I domain. The protein forms a ternary complex with insulin-like growth factor acid-labile subunit (IGFALS) and either insulin-like growth factor (IGF) I or II. In this form, it circulates in the plasma, prolonging the half-life of IGFs and altering their interaction with cell surface receptors.

In Figure 91 the expression ratios after Rifampicin induction are given for several Cytochrome P450s and Glutathione S-transferases which are partially known to be induced by xenobiotic substances. While CYP1A1, CYP1A2, CYP2C18, CYP4F3, CYP1B1, CYP2D6 showed only weak or almost no change of the expression ratio between induced and control cells, a strong change was observed for CYP3A4, CYP2B6, CYP2A6, CYP2C8, CYP4A11 and CYP2E1. While CYP4A11 (1.7-fold) and CYP2E1 (3-fold) were remarkably down-regulated after 24 as well as after 72 h of induction, the other differentially expressed CYPs showed an up-regulation. CYP3A4, CYP2B6 and CYP2C8 were up-regulated 2-fold after 6 h. CYP3A4 and CYP2B6 were up to 4-fold induced in Rifampicin treated cells compared to controls after 24 h. After 72 h this induction remained for CYP3A4 while the ratio for CYP2B6 declined to 3.5-fold induction. The expression of CYP2C8 increased only slightly after 24 h of induction to 2.2-fold, while after 72 h it was 4.7-fold induced. CYP2A6 showed a complete different expression profile. Starting with a ratio of about 1 after 6 h the expression increased in Rifampicin treated cells and the ratio after 24 h was determined with a 2.3-fold induction. After 72 h no induction was monitored for CYP2A6 the ratio was again measured to be around 1. Concerning the expression change of glutathione reductase (GSR), glutathione synthetase (GSS) and glutathione S-transferases (GST) only the expression of GSTA1 and GSTM3 showed an altering profile over time. All other given genes in this graph were not differentially expressed in Rifampicin treated cells.

Glutathione S-transferase A1 (GSTA1) which had a very low expression ratio after 6 h was strongly induced after 24 h (4-fold). After 72 hours a 1.8-fold down-regulation was detected. The coded enzymes function in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione. The genes encoding these enzymes are known to be highly polymorphic. These genetic variations can change an individual's susceptibility to carcinogens and toxins as well as affect the toxicity and efficacy of some drugs. At present, eight distinct classes of the soluble cytoplasmic mammalian glutathione S-transferases have been identified: alpha, kappa, mu, omega, pi, sigma, theta and zeta. This gene *gstA1* encodes a glutathione S-transferase belonging to the alpha class. The alpha class genes, located in a cluster mapped to chromosome 6, are the most abundantly expressed glutathione S-transferases in liver. In addition to metabolizing bilirubin and certain anti-cancer drugs in the liver, the alpha class of these enzymes exhibit glutathione peroxidase activity thereby protecting the cells from reactive oxygen species and the products of peroxidation.

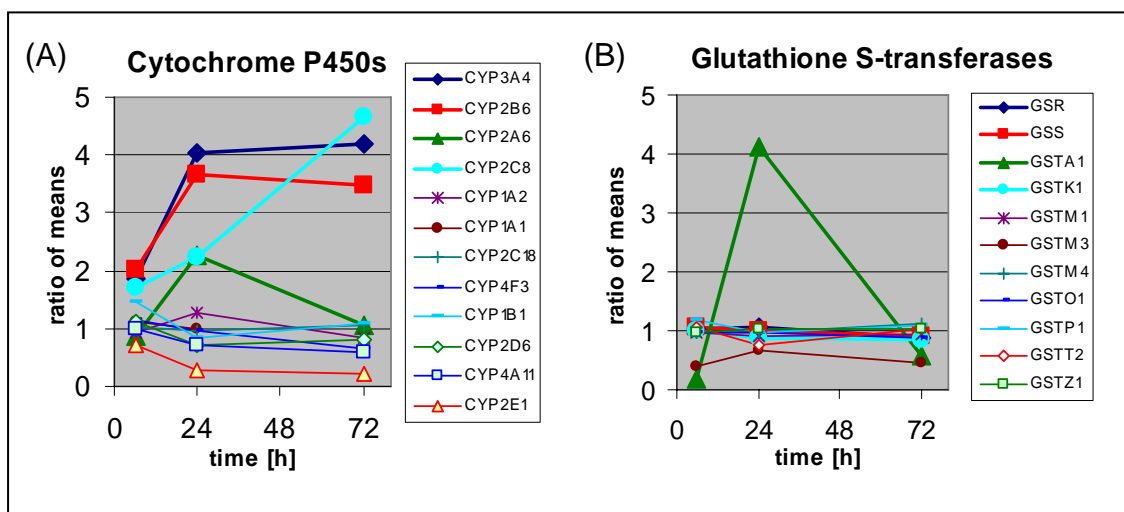


Figure 91: Expression profiles of (A) Cytochrome P450 monooxygenases and (B) Glutathione S-transferases. In part (A) the expression ratios for several CYPs are plotted over time. In part (B) the expression ratios between DMSO control samples and Rifampicin treated cells after 6, 24 and 72 hours are given for several Glutathione S-transferases.

It is interesting to notice that the expression of GGS is not affected by the induction. Either the amount of available glutathione has not to be increased due to induction or the capacity of the present GSS is high enough to fulfil its task to eliminate reactive oxygen

species. Glutathione is important for a variety of biological functions, including protection of cells from oxidative damage by free radicals, detoxification of xenobiotics, and membrane transport. The protein encoded by the gene *gss* functions as a homodimer to catalyze the second step of glutathione biosynthesis, which is the ATP-dependent conversion of gamma-L-glutamyl-L-cysteine to glutathione.

In Figure 92 several more genes involved in drug metabolism are summarized. Part one shows epoxide hydrolases (EPHX1 and 2) which play an important role in both the activation and detoxification of exogenous chemicals such as polycyclic aromatic hydrocarbons. Furthermore flavin containing monooxygenases (FMO4 and 5) are given. Flavin-containing monooxygenases are NADPH-dependent flavoenzymes that catalyze the oxidation of nucleophilic heteroatom centres in drugs, pesticides, and xenobiotics. In addition monoamine oxidases (MAOA and B) and N-acetyltransferase 1 (NAT1) are shown. MAOA is an enzyme that degrades amine neurotransmitters, such as dopamine, norepinephrine, and serotonin and other compounds. MAOB also belonging to the flavin monoamine oxidase family catalyzes the oxidative deamination of biogenic and xenobiotic amines and plays, as MAOA, an important role in the metabolism of neuroactive and vasoactive amines in the central nervous system and peripheral tissues.

EPHX1 was found to be up-regulated over the whole experiment in Rifampicin induced cells. After 6 h an almost 2-fold and after 24 and 72 h a 1.6-fold increase in expression was detected. MAOB was 2.3-fold induced after 6 h and strongly down-regulated after 24 h (0.3-fold). After 72 hours no differentially expression was observed.

All other genes investigated in this group (drug metabolism I) showed only a slight change of their expression profile over time.

Part two of Figure 92 displays genes from the phase II of xenobiotic metabolism. The expression ratio of three sulfotransferases (SULT1A1, SULT1E1, and SULT2A1) which catalyze the sulfate conjugation of many drugs, and xenobiotic compounds and two UDP glucuronosyltransferases (UGT1A1 and UGT2B15) are shown. UGTs are enzymes of the glucuronidation pathway that transforms small lipophilic molecules, such as steroids, bilirubin, hormones, and drugs, into water-soluble, excretable metabolites.

The overall expression change for these genes of phase II of xenobiotic metabolism was low. SULT1A1 and SULT1E1 were down-regulated while SULT2A1 showed a slight up-regulation after 24 h. UGT1A1 was up-regulated after 24 h (1.3-fold) and maintained this up-regulated state till 72 h after induction. UGT2B15 was first slightly down-regulated (1.4-fold) after 24 h and then slightly up-regulated after 72 h (1.2-fold) of induction.

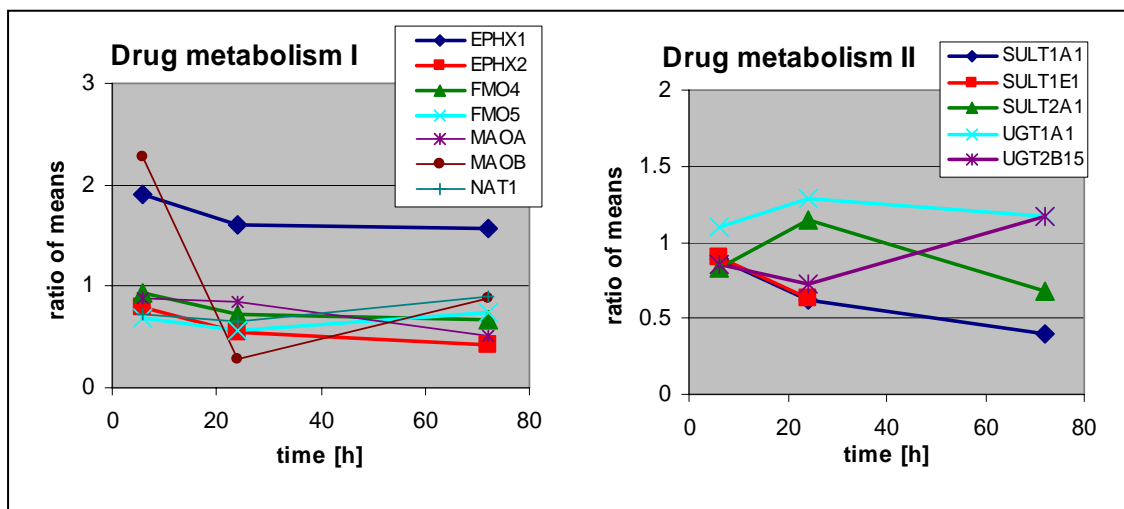


Figure 92: Expression profiles of genes involved in drug metabolism. The expression ratios between DMSO control samples and Rifampicin treated cells after 6, 24 and 72 hours are given for EPHX1, EPHX2, FMO4, FMO5, MAOA, MAOB, NAT1, SULT1A1, SULT1E1, SULT2A1, UGT1A1 and UGT2B15.

In Figure 93 expression ratios for genes having an impact on transcription are shown. Aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor involved in the regulation of biological response to planar aromatic hydrocarbons. As this receptor has been shown to regulate xenobiotic-metabolising enzymes such as cytochrome P450 it is of special interest in this study. Its ligands included a variety of aromatic hydrocarbons. Like its target genes (CYPs) this receptor shows a slight up-regulation (1.1-, 1.2-, 1.8-fold) after 6, 24 and 72 hours of induction. The also well known and characterized nuclear receptors CAR (113) and PXR (112) showed a down-regulation after 24 h. PXR had a ratio of 0.9 after 6 h and declined to 0.7 and 0.6 after 24 and 72 h, resulting in a 1.7-fold down-regulation. CAR showed an even stronger down-

regulation. This receptor was down-regulated 3.3-fold (ratio 0.3) after 24 h and was not detectable after 72 h.

CCAAT/enhancer binding protein (C/EBP), alpha (CEBPA) is a transcription factor which can bind as a homodimer to certain promoters and enhancers. The encoded protein has been shown to bind to the promoter and modulate the expression of several genes. The encoded protein can interact with CDK2 and CDK4, thereby inhibiting these kinases and causing growth arrest in cultured cells. CEBPA was slightly down-regulated after 24 h and showed no differential expression after 24 and 72 hours.

Glucocorticoid receptor DNA binding factor 1 (GRLF1) associates with the promoter region of the glucocorticoid receptor gene (*hgr*). It is a repressor of glucocorticoid receptor transcription. The GRLF1 enhances the homologous down-regulation of wild-type *hgr* gene expression. Biochemical analysis suggests that GRLF1 interaction is sequence specific and that transcriptional efficacy of GRLF1 is regulated through its interaction with specific sequence motif. The level of expression is regulated by glucocorticoids. After 24 h a down-regulation was observed for GRLF1 from 1.2-fold induction after 6 h to a 1.4-fold down-regulation. GRLF1 was higher expressed in Rifampicin treated cells compared to control cells with a ratio of 1.3-fold after 72 h of induction.

A similar expression profile as for GRLF1 was observed for HNF4 α . Hepatocyte nuclear factor 4, alpha is a nuclear transcription factor which binds DNA as a homodimer. It controls the expression of several genes, including hepatocyte nuclear factor 1 alpha, a transcription factor which regulates the expression of several hepatic genes. From equal expression in Rifampicin and DMSO cells after 6 h HNF4 α was first down-regulated after 24 h (1.8-fold). Its expression was increased again after 72 h to no differential expression between induced and control samples.

Nuclear factor kappa light polypeptide gene enhancer (NF κ B1) which is a transcription regulator that is activated by various intra- and extra-cellular stimuli such as cytokines, oxidant-free radicals, ultraviolet irradiation, and bacterial or viral products showed a slightly elevated (1.2) but not changing expression profile. If activated NF κ B translocates into the nucleus and stimulates the expression of genes involved in a wide variety of biological functions.

Peroxisome proliferator-activated receptors (PPAR α , δ , γ) belong to the steroid hormone receptor superfamily. PPARs affect the expression of target genes involved in cell proliferation, cell differentiation and in immune and inflammation responses. Three closely related subtypes (alpha, beta/delta, and gamma) have been identified. Subtype PPAR-alpha is a nuclear transcription factor. PPAR-delta is a potent inhibitor of ligand-induced transcription activity of PPAR-alpha and PPAR-gamma. It may function as an integrator of transcription repression and nuclear receptor signalling. PPARs showed a down-regulation after 24 h of induction. PPAR-delta persisted at a not differentially expressed state while PPAR-alpha and -gamma increased slightly after 72 hours. Retinoic acid receptor, alpha (RAR α) showed no differential expression over time.

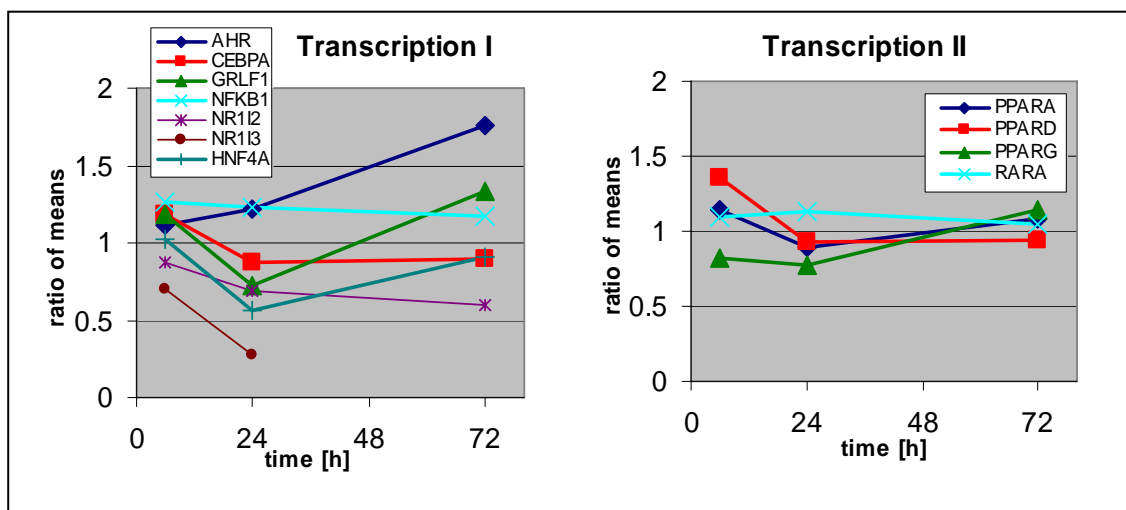


Figure 93: Expression profiles of genes involved in regulation of transcription. The expression ratios between DMSO control samples and Rifampicin treated cells after 6, 24 and 72 hours are given for AHR, CEBPA, GRLF1, NFKB1, NR112, NR113, HNF4 α , PPARA, PPARD, PPARG, RARA.

In Figure 94 genes involved in phase III of xenobiotic metabolism are shown in part (A) and part (B) shows those involved in lipid metabolism. ATP-binding cassette, sub-family G, member 2 (ABCG2) functions as a xenobiotic transporter which may play a major role in multi-drug resistance. However ABCG2 showed no differential expression in the treated cells compared to controls over time. Solute carrier family (SLCs) polyspecific organic cation transporters found in the liver, kidney, intestine, and other organs are critical for elimination of many endogenous small organic cations as

well as a wide array of drugs and environmental toxins. Nevertheless SLC15A1, SLC22A7, SLC22A9 and SLC22A1 were down-regulated in Rifampicin treated cells over time.

The genes selected to be involved in lipid metabolism (part B of Figure 94) showed little change in expression level. Acyl-coenzyme A oxidase 1 (ACOX1) the first enzyme of the fatty acid beta-oxidation pathway, which catalyzes the desaturation of acyl-CoAs to 2-trans-enoyl-CoAs showed a slight down-regulation of the expression in Rifampicin induced cells after 24 h. A similar observation was done for apolipoprotein E (APOE). APOE, a main apoprotein of the chylomicron, binds to a specific receptor on liver cells and peripheral cells. APOE is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. Enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase (EHHADH) was down-regulated after 72 h. Prostaglandin E synthase 2 (PTGES2) was down-regulated in Rifampicin treated cells compared to DMSO control cells in all points of time.

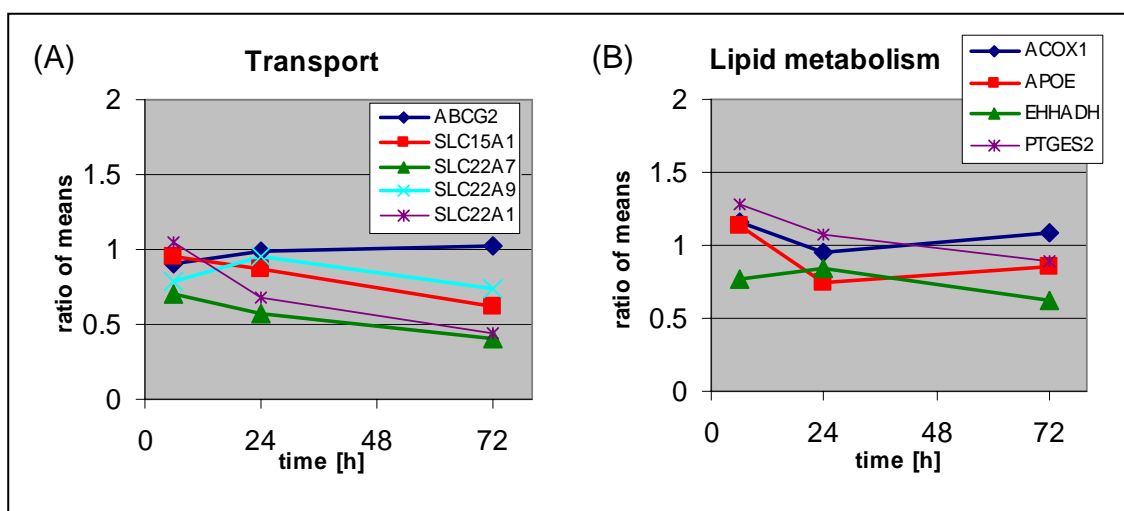


Figure 94: Expression profiles of genes involved in (A) transport and (B) lipid metabolism. In part (A) the expression ratios for genes involved in transport (ABCG2, SLC15A1, 22A7, 22A9, 22A1) are plotted over time. In part (B) the expression ratios between DMSO control samples and Rifampicin treated cells after 6, 24 and 72 hours are given for genes playing a role in lipid metabolism (ACOX1, APOE, EHHADH, PTGES2).

For normalization of expression data and as a control for viability of hepatocyte cells the expression profiles of 13 genes, considered as housekeeping genes which should not alter their expression during induction. The expression profiles are shown in Figure 95.

Except for malate dehydrogenase 1 (MDH1) and serine dehydratase (SDS) which showed a 1.5-fold differential expression the other housekeeping genes were constant and similar expressed in induced and control samples.

MDH1 catalyzes the reversible oxidation of malate to oxaloacetate, utilizing the NAD/NADH cofactor system in the citric acid cycle. It showed an up-regulation in induced cells after 24 h (1.7-fold). After 72 h no differential expression was observed. SDS is found predominantly in the liver and is involved in metabolizing serine and glycine. L-serine dehydratase converts L-serine to pyruvate and ammonia and requires pyridoxal phosphate as a cofactor. The encoded protein can also metabolize threonine to NH_4^+ and 2-ketobutyrate. It was induced after 6 h and then down-regulated after 24 h to a not differentially expressed state.

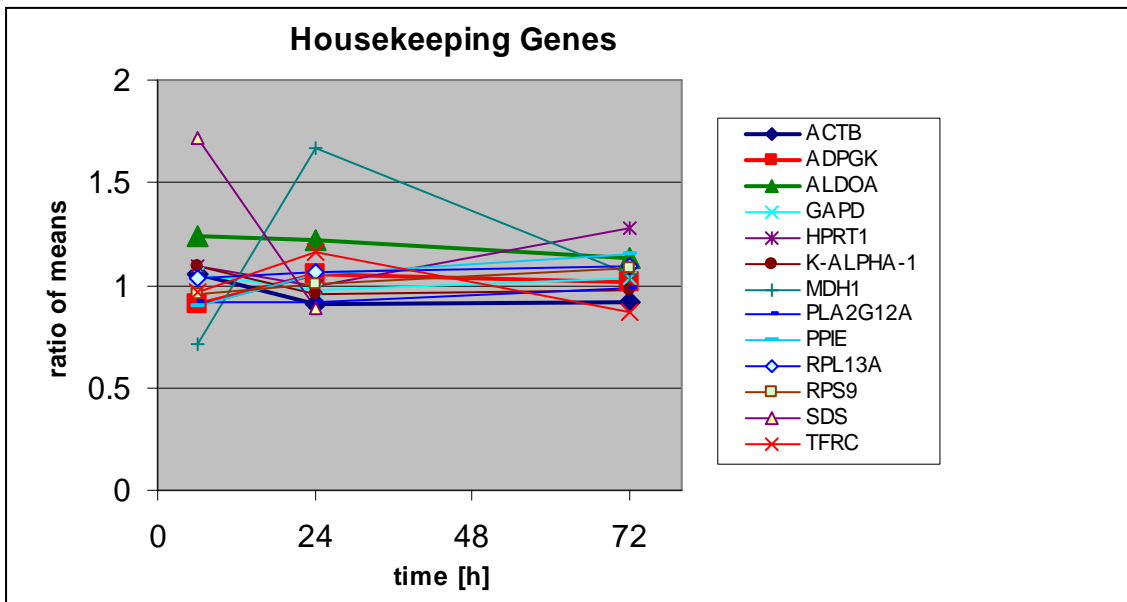


Figure 95: Expression ratios of housekeeping genes in primary human hepatocytes after Rifampicin induction in induced versus control cells. The expression was measured after 6, 24 and 72 h and the ratios between control- and induced cells are plotted over time.

3.8.3 Reconstruction of Networks Using a Combined Literature Mining and Microarray Analysing Approach (LMMA)

To add to expression data a functional context a literature mining and microarray analysing approach (LMMA) was applied. In a first step the significant differentially expressed genes were determined for samples 24 h after induction. This point of measurement was chosen as previous analyzes had revealed the strongest effect of Rifampicin induction to be after 24 hours of induction. A graphical overview of the obtained results is given in Figure 96. The observed ratio is plotted versus the expected ratio. Significant features can be found below (down-regulated) or above (up-regulated) the red lines which mark the border between significant and not significant observations. 725 significant up-regulated features and 757 down-regulated features were detected.

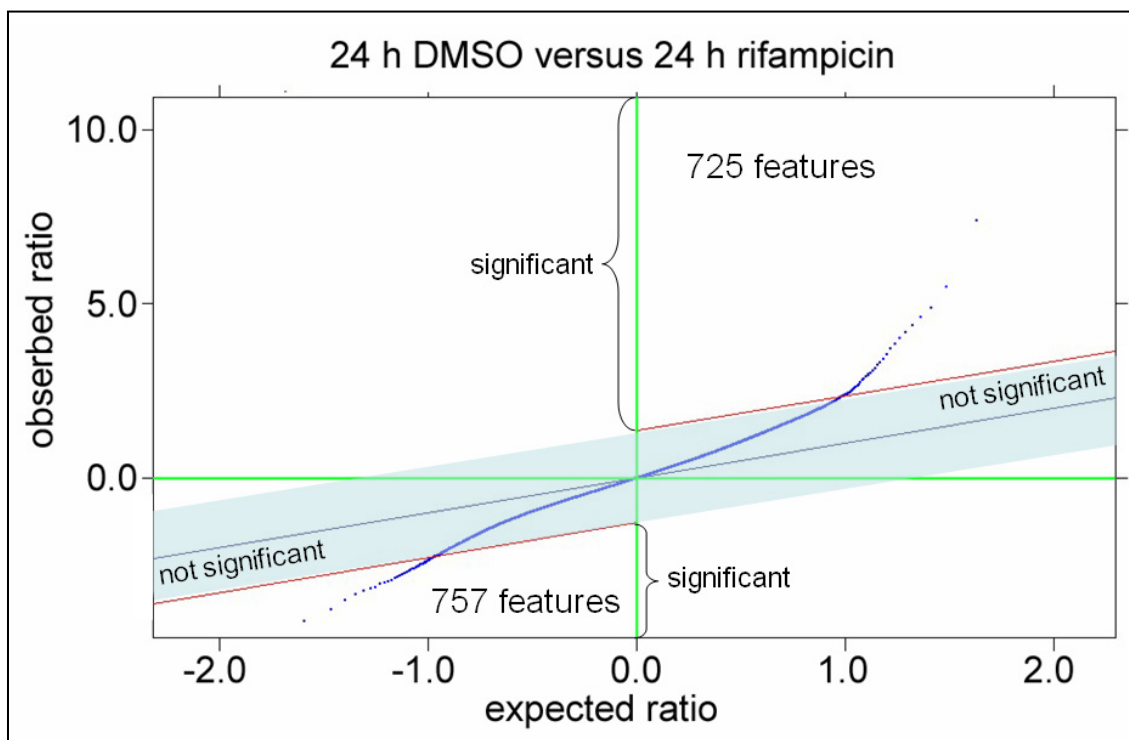


Figure 96: Significant differentially expressed features on the full genome arrays from Affymetrix. After 24 h of induction with Rifampicin 757 features on the Affymetrix arrays were determined to be significantly down-regulated while 725 were regarded as significantly up-regulated. The observed ratio was plotted against the expected ratio to obtain a measure of significance. For details of the method see Tusher *et al.*, 2001.

The detected significantly differentially expressed genes were grouped using the following go annotation categories based on the biological process (Figure 97) and molecular function (Figure 98). For each observation a z-score was calculated. The higher the z-score the more important is the group / function for the total change.

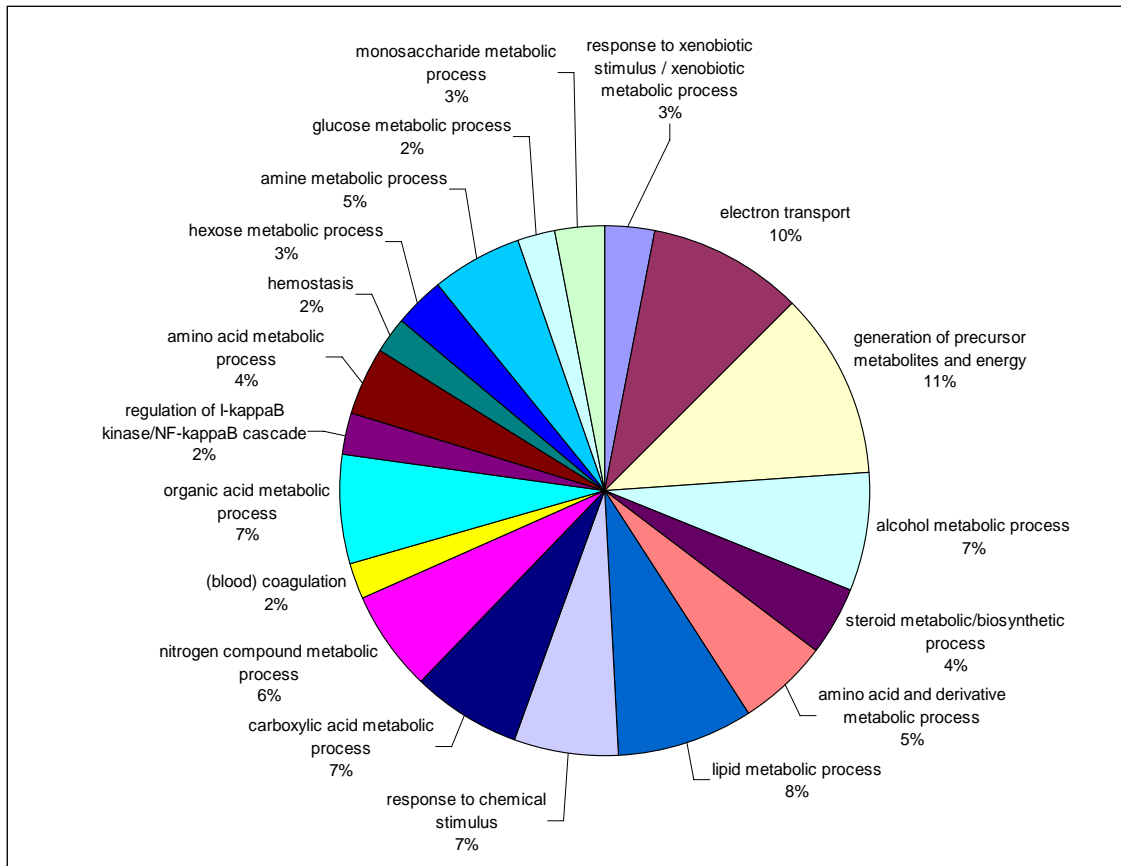


Figure 97: Most affected biological processes after induction of hepatocyte cells with Rifampicin. Significantly differentially expressed features were sorted based on the Go annotation categories.

The most affected biological processes were: the xenobiotic metabolic process (GO: 0006805) / response to xenobiotic stimulus (GO: 0009410) where five genes of 34 in the total group were found to be altered yielding in a z-score of 9.2. In total this group accounted for three percent of the total change of biological processes as shown in Figure 97. The second highest overrepresented biological process was electron transport (GO: 0006118). 316 genes are annotated with this function and 16 were found altered in this experiment accounting for 10 % of the overall change. The z-score for this observation was 8.7. The third process was generation of precursor metabolites and

energy (GO: 0006091) with 19 of 519 genes and a z-score of 7.6 followed by alcohol metabolic process (GO: 0006066) and steroid metabolic process (GO: 0008202). All other minor important, but still significant (Z -score > 4) biological processes are displayed in Figure 97.

The most significant changed molecular functions after 24 hours of Rifampicin induction were cadmium ion binding, oxidoreductase activity as well as unspecific monooxygenase activity, alcohol dehydrogenase activity, monooxygenase activity and heme binding. The percentage of the total change on expression level for all significant changed cellular functions are given in Figure 98.

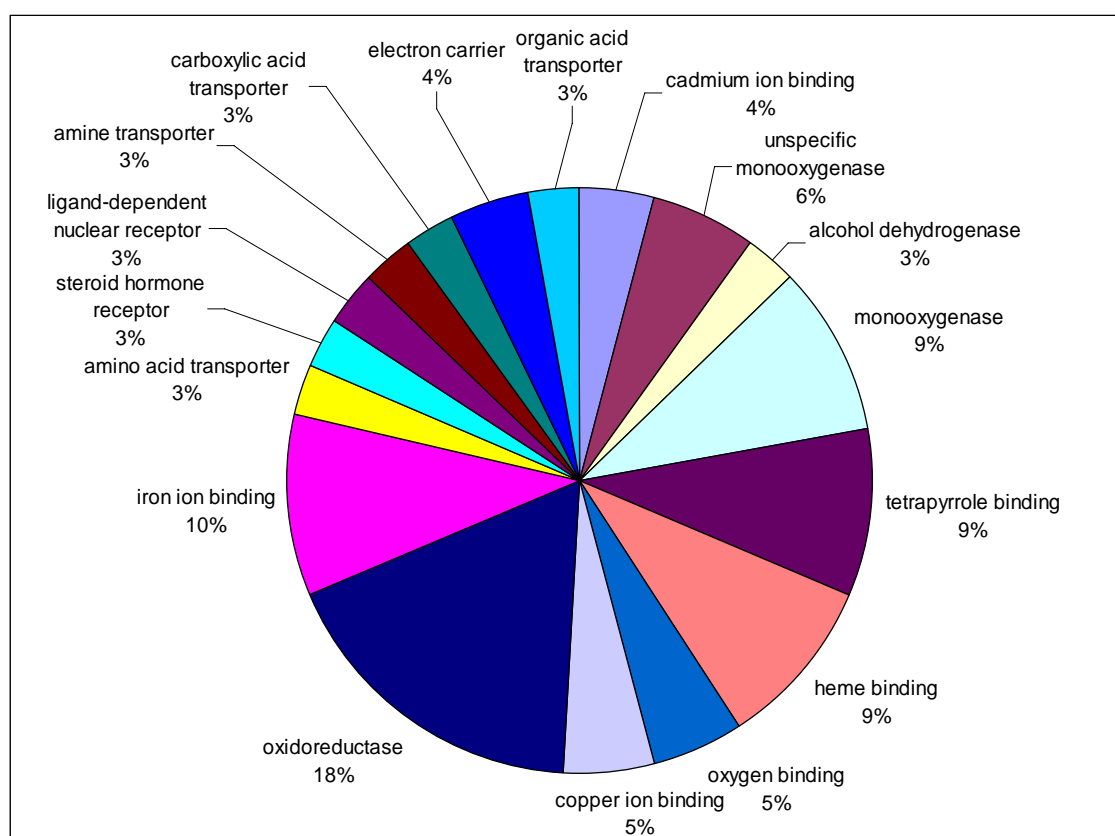


Figure 98: Most affected molecular functions after induction of hepatocyte cells with Rifampicin. Significantly differentially expressed features were sorted based on the Go annotation categories.

In the cadmium ion binding activity (GO:0046870) six of the eight genes of this category showed an altered expression accounting for 4 % of the change. The z-score was determined with 23.8. In the oxidoreductase activity group (GO:0016712) eleven

of 31 genes were differentially expressed yielding in a z-score of 21.9. 18% of the change in expression occurred in this group.

Eight of the 25 genes in the unspecific monooxygenase activity group (GO:0050381) were found to be differentially expressed. 6 % of the expression change occurred in this category. The calculated z-score for this observation was 17.7. For the alcohol dehydrogenase activity (GO:0004022) four of the ten genes were detected to be differentially expressed. The z-score was 14.0 and the percentage on the significant gene expression change was 3 %.

For the monooxygenase activity (GO:0004497) 13 of 98 genes were differentially expressed giving a z-score of 14.0 and accounting for 9 % of the expression change. In the heme binding group (GO:0020037) consisting of 103 genes, 13 genes were differentially expressed with a z-score of 13.6 for this observation. Genes in the heme binding group account for 9 % of the significantly overrepresented molecular functions. All other minor important, but still significant (Z-score > 4) functional categories are displayed in Figure 98.

To extract significant information for network reconstruction from the expression level of single probes of Affymetrix GeneChip microarrays ChipInspector 1.35 from Genomatix was used. The resulting lists of significantly regulated genes from the experiment were used to reconstruct the regulatory network of CYP3A4.

To avoid wrong annotation previous annotations of the single oligonucleotide probes were disregarded together with the grouping of the probes in probe sets. Mismatch probes were disregarded as well. The sequence of each single probe was mapped against the current human genome and against Genomatix' database of transcripts. Only probes that met quality criteria such as uniqueness in the genome and mismatch proof were used for the analysis. More than 75 % of the perfect match probes could be used to calculate the statistics for the Human Genome U133 Plus 2.0 array. 495.007 perfect match probes fulfil the quality criteria and were used for the assessment of the expression of 48.811 human transcripts. The normalization was done using a linear total intensity normalization algorithm. For statistical evaluation a significance test was performed at the single probe level. This was done via a standard permutation t-test, similar to SAM Tusher *et al.*, 2001. A set of probe-specific t-tests was conducted to identify significantly regulated single probes. Each probe obtained a score on the basis

of its fold change relative to the standard deviation of repeated measurements for this probe. Probes with scores higher than a certain threshold are deemed significant. This threshold is the so called delta or d-value. Decreasing the delta value yields more significant features at the cost of increasing likeliness to find false positives features. This means increasing the delta value diminishes the group of detected probes, but the resulting group has a higher stringency, i.e. less danger of including falsely detected probes. Permutations of the data were then used to estimate the percentage of probes identified by chance at this delta value, this is the false discovery rate (FDR). The FDR is a confidence value giving an idea of how many probes in the result group are possibly falsely selected. A low FDR denotes a high stringency, in other words less probes are recognized as being statistically significant. The dependency between found differentially expressed features and FDR as well as delta value is given in Table 48.

Table 48: Dependency of significant differentially expressed features and statistical parameters (delta-value and FDR). With a lower delta value for up-regulated and a higher for down-regulated features the total number of observed differentially expression increases. With a higher FDR the number of detected features but also the number of falsely detected features increases.

delta value	FDR / falsely called features	Number of significant features	
		up-regulated	down-regulated
- 2.7 / 2.7	0 / 0	147	9
- 2.5 / 2.5	0 / 0	177	20
- 2.0 / 2.0	0 / 0	289	75
- 1.5 / 1.5	0 / 0	530	442
- 1.3 / 1.3	0 / 0	813	757
- 1.2 / 1.2	0.1 / 0.6	1100	1015
- 1.0 / 1.0	0.2 / 3.0	2017	1786
- 0.8 / 0.8	0.5 / 23.0	5091	3685

For the up-regulated probes the d-value 1.35 was selected. In this case 725 features fulfilled the quality criteria with a false discovery rate of zero. For the down-regulated features a delta of -1.3 was chosen yielding in 757 significant features with a false discovery rate of zero (Figure 96).

These features were correlated with 341 transcripts based on "minimum probe coverage" of three. Probe coverage is the number of significant probes needed to make this transcript appear in the result list. It had to be further taken into account that not all

probes were placed in coding regions of the gene. The signals of probes in non coding regions were excluded.

For each transcript, the ratio between experiment and reference is displayed as the average expression ratio of all significant probes that match this transcript. The correlation between probe coverage, ratio and number of significantly expressed features is given in Table 49.

Table 49: Correlation between probe coverage, expression ratio and number of differentially expressed transcripts. With a higher degree of probe coverage (shown in column one) the number of identified transcripts decreases. An increased expression ratio threshold (column two) also reduces the number of transcripts using the same probe coverage.

minimum probe coverage	minimum expression ratio	number of transcripts
6	0	143
5	0	201
4	0	258
3	0	341
3	0.5	341
3	1.0	65
3	1.5	8

For further analyses a so called "probe coverage" of three was selected with a minimum expression ratio of 0.5. This resulted in 341 significant differentially expressed features. These 341 features were then mapped to their corresponding genes. Finally 156 unique identifiers could be mapped to a gene. The gene list is shown in Table 65 in the appendix. The 20 most up- and down-regulated genes are summarized in Table 50 and Table 51. Within the 20 most up-regulated genes six CYP genes were identified. This analysis showed CYP3A4 which is the centre of the following network reconstruction to be the most up-regulated (2.2-fold) gene after 24 h of induction. Metallothionein 1F (1.8 fold) and aminolevulinate, delta-, synthase 1 (1.6) showed a strong induction (> 1.5 fold). Six further members from the metallothionein group were found in the group of the 20 most up-regulated genes. As far as transporters were concerned ATP-binding cassette, sub-family B (MDR/TAP), member 1 (ABCB1) was found to be up-regulated 1.2 fold.

Table 50: Top 20 genes showing the highest significant up-regulation after 24 hours of Rifampicin induction based on the statistical test from Tusher et al., 2001.

No	Symbol	Gene Name	Ratio
1	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	2.2
2	MT1F	metallothionein 1F	1.8
3	CYP3A43	cytochrome P450, family 3, subfamily A, polypeptide 43	1.7
4	ALAS1	aminolevulinate, delta-, synthase 1	1.6
5	MT1X	metallothionein 1X	1.4
6	CYP3A7	cytochrome P450, family 3, subfamily A, polypeptide 7	1.4
7	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 peg. 1	1.4
8	AGXT2L1	alanine-glyoxylate aminotransferase 2-like 1	1.3
9	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	1.3
10	MALAT1	metastasis associated lung adenocarcinoma transcript 1	1.3
11	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	1.2
12	MT1E	metallothionein 1E	1.2
13	CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8	1.2
14	MT1M	metallothionein 1M	1.2
15	SEC14L4	SEC14-like 4 (<i>S. cerevisiae</i>)	1.1
16	OSTbeta	organic solute transporter beta	1.1
17	MT1G	metallothionein 1G	1.1
18	MT2A	metallothionein 2A	1.1
19	PRODH2	proline dehydrogenase (oxidase) 2	1.1
20	MT1H	metallothionein 1H	1.1

The strongest down-regulation was found for several genes involved in metabolism of alcohol: alcohol dehydrogenase IB (class I), beta polypeptide (ADH1b -1.6), alcohol dehydrogenase 4 (class II), pi polypeptide (ADH4 -1.2), alcohol dehydrogenase 1A (class I), alpha polypeptide (ADH1A -1.1), alcohol dehydrogenase 6 (class V) (ADH6 -1.0) and cytochrome P450, family 2, subfamily E, polypeptide 1 (CYP2E1 -1.1). Further down-regulated genes, already known to be involved in detoxification were: cytochrome P450, family 7, subfamily A, polypeptide 1 (CYP7A1 -1.4), nuclear receptor subfamily 1, group I, member 3 (NR1I3, CAR -1.2), solute carrier family 38, member 4 (SLC38A4 -0.9) and solute carrier family 6 (neurotransmitter transporter), member 1 (SLC6A1 -0.9).

Table 51: Top 20 genes showing the highest significant down-regulation after 24 hours of Rifampicin induction based on the statistical test from Tusher et al., 2001.

No	Symbol	Gene Name	Ratio
1	ADH1B	alcohol dehydrogenase IB (class I), beta polypeptide	-1.6
2	CYP7A1	cytochrome P450, family 7, subfamily A, polypeptide 1	-1.4
3	H19	H19, imprinted maternally expressed untranslated mRNA	-1.3
4	TAT	tyrosine aminotransferase	-1.3
5	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2	-1.3
6	NR1I3	nuclear receptor subfamily 1, group I, member 3	-1.2
7	ADH4	alcohol dehydrogenase 4 (class II), pi polypeptide	-1.2
8	CYP2E1	cytochrome P450, family 2, subfamily E, polypeptide 1	-1.1
9	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	-1.1
10	ADH1A	alcohol dehydrogenase 1A (class I), alpha polypeptide	-1.1
11	MAT2A	methionine adenosyltransferase II, alpha	-1.0
12	INHBE	inhibin, beta E	-1.0
13	SERPIN	serpin peptidase inhibitor, clade A (antitrypsin), member 10	-1.0
14	ADH6	alcohol dehydrogenase 6 (class V)	-1.0
15	SLC38A4	solute carrier family 38, member 4	-0.9
16	RDH16	retinol dehydrogenase 16 (all-trans)	-0.9
17	IFIT2	interferon-induced protein with tetratricopeptide repeats 2	-0.9
18	SLC6A1	solute carrier family 6 (neurotransmitter transporter), member 1	-0.9
19	PDK4	pyruvate dehydrogenase kinase, isozyme 4	-0.9
20	PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)	-0.9

In a first approach only genes with the highest fold change were used for a literature based network reconstruction. Literature (from public and Genomatix databases) and microarray gene-expression data (from this study) were integrated. In a combined literature mining and microarray analysis (LMMA) approach a gene network of CYP3A4 interaction was constructed using the literature-based co-occurrence method. Li *et al.*, 2006 observed that the LMMA-based network is more reliable than the co-occurrence network based on multiple levels of Kyoto Encyclopaedia of Genes and Genomes (KEGG) -gene, -ontology and -pathways.

A small network taking only the strong differentially expressed genes into account is shown in Figure 99. The interaction between CAR (NR1I3) and CYP3A4 is detected. The interaction of CYP3A4 with PXR (NR1I2) is not shown as PXR showed only a small change in expression upon stimulating with Rifampicin. Therefore PXR does not show up although it is crucial for inducing CYP3A4.

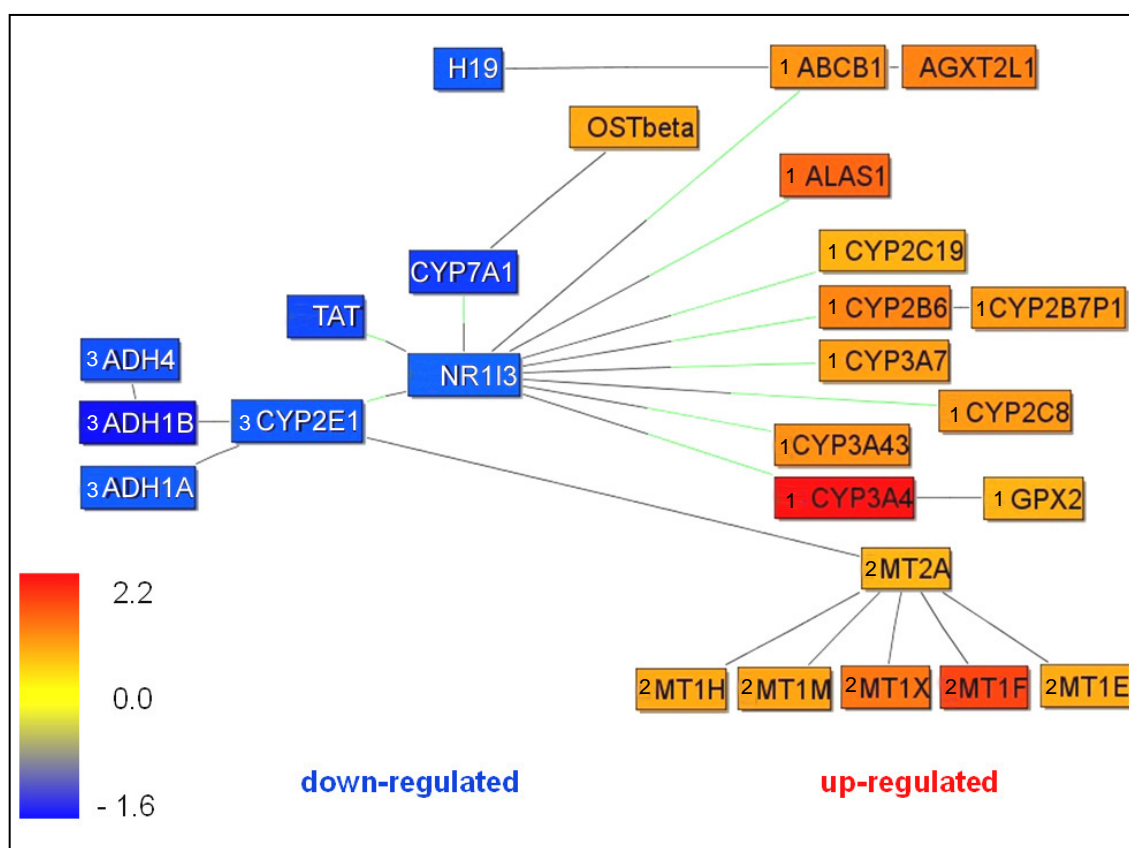


Figure 99: Network reconstruction based on co-occurrence in literature and microarray data. In this small network only the genes with the highest change in expression were taken into account. The colour code indicates the expression level of the genes. Blue stands for down-regulation, while red indicates up-regulation. Three functional clusters could be identified (1) CYP-cluster and detoxification of xenobiotics, (2) metallothionein-cluster protecting against metal toxicity and oxidative stress and (3) Omega oxidation of fatty acids and alcohol metabolism (ADHs and CYP2E1). The black lines give not curated annotations of interaction. The black-green lines connect genes where the transcription factor binding side match into the target promoter of the linked gene.

All 156 identified genes were used as the framework for a larger network reconstruction. A part of this larger network, built around the target of our investigations CYP3A4 is given in Figure 100. If the network is extended to genes showing a lower change in expression, also PXR (NR1I2) is shown as interaction partner with CYP3A4. Further connections to RXR α and HNF4 α are given. It is interesting to notice that the whole regulatory systems seems to be down-regulated in primary hepatocyte cells induced with Rifampicin compared to not induced cells. CAR (NR1I3), PXR (NR1I2) and HNF4 α were all detected to be down-regulated, indicated by the blue colour in Figure 100. However targets of PXR (NR1I2) like CYP3A4 and

ALAS1 were strongly up-regulated. As detected previously several members of the CYP family are up-regulated as well as ABCB1 and several metallothionein proteins.

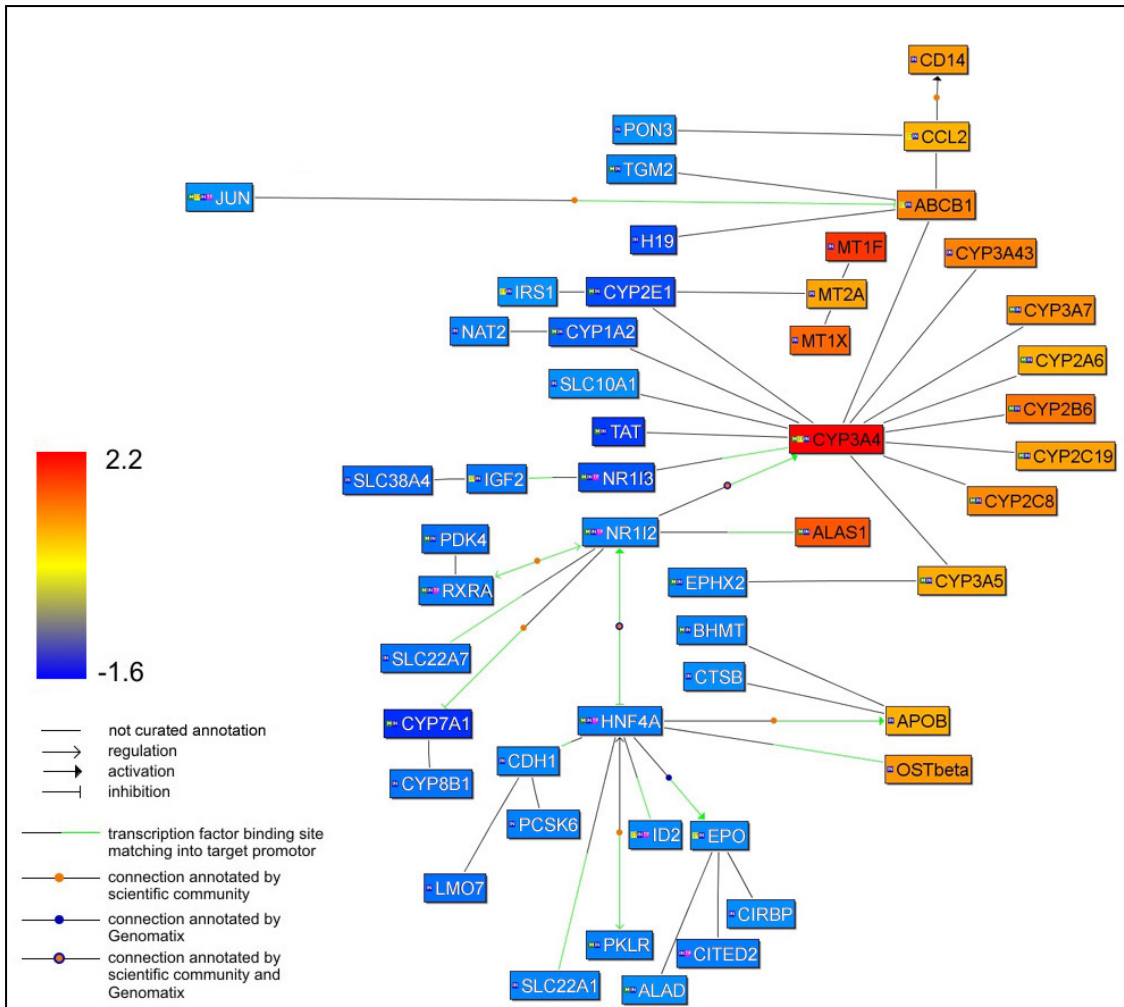


Figure 100: Network reconstruction based on co-occurrence in literature and microarray data. This interaction network shows the connection between the central genes CYP3A4, PXR, CAR, HNF4 α and their periphery. The expression levels are colour coded. Down-regulated genes are given in blue, while up-regulated ones are shown in red. The explanation of the connecting lines and arrows can be found in the legend.

The expression of genes involved in the transcriptional regulation of CYP3A4 was monitored. An overview of the regulatory mechanism of CYP3A4 induction by Rifampicin, based on Li and Chiang, 2006, is given in Figure 101.

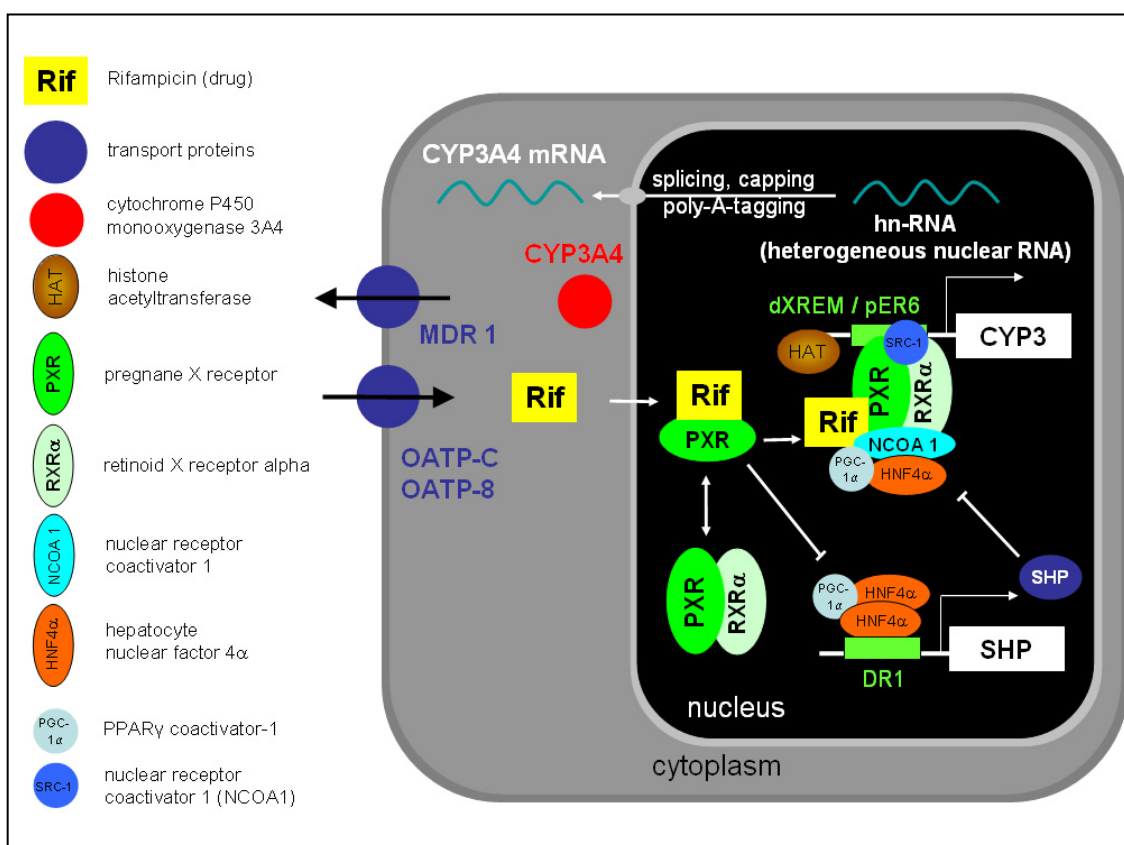


Figure 101: Proposed model of regulation of expression of CYP3A4 in hepatocytes. A complex formed by PXR/RXR α /Rifampicin/NCOA 1/HNF4 α /PGC-1 α (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha) and SRC-1 (NCOA1, p160/steroid receptor coactivator 1) binds to the regulatory elements of the human CYP3A4 promoter. Rifampicin binds to PXR and induces the interaction with HNF4 α . HNF4 α is co-activated by PGC-1 α and SRC-1. SHP is able to inhibit CYP3A4 gene transcription by blocking PXR and HNF4 α interaction. SHP transcription is induced by binding of HNF4 α to the SHP promoter. Rifampicin activated PXR inhibits SHP transcription. SHP expression is reduced and CYP3A4 transcription is enhanced (Li and Chiang, 2006).

The protein peroxisome proliferator-activated receptor gamma, co-activator 1 alpha (PPARGC1A, PGC-1 α) is a transcriptional co-activator that regulates the genes involved in energy metabolism. This protein interacts with PPARgamma, which permits the interaction of this protein with multiple transcription factors. It seems to play an important role in forming the Rif/PXR/RXR/HNF4 α complex activating CYP3A4 expression. Furthermore this protein can interact with, and regulate the activities of, cAMP response element binding protein (CREB) and nuclear respiratory factors (NRFs). It provides a direct link between external physiological stimuli and the regulation of mitochondrial biogenesis, and is a major factor that regulates reactions of energy- and carbon-metabolism. This protein may be also involved in controlling blood

pressure, regulation of cellular cholesterol homeostasis, and the development of obesity. It was similarly expressed in culture HH26 and HH44 with a slight down-regulation after 24 h. After 72 h PGC1a reached the initial expression state. In culture HH27 it was down-regulated after 6 h and increased expression to a not differentially expressed state after 24 and 72 hours.

The hepatocyte nuclear factor 4, alpha (HNF4 α) is a nuclear transcription factor which binds DNA as a homodimer. The encoded protein controls the expression of several genes. This gene may play a role in development of the liver, kidney, and intestines. HNF4 α showed a similar expression profile for all three cultures. It was highly over-expressed in culture HH27 after 6 h while in the other two cultures from donor 26 and 44 no differentially expression was observed at this point of time. After 24 h the expression of HNF4 α was declining to about 0.5 (2-fold down-regulation). This down-regulation was maintained after 72 h.

Nuclear receptor co-activator 1 (SRC-1, NCOA 1) acts as a transcriptional co-activator for steroid- and nuclear hormone receptors. The product of this gene binds nuclear receptors directly and stimulates the transcriptional activities in a hormone-dependent fashion. SRC-1 showed almost no change in expression over time between Rifampicin treated- and control cells. The expression pattern of SRC-1 was different in HH27 cell culture compared to HH26 and HH44. While in HH27 after a slight induction at 24 h the increase of expression was going down again after 72 h in culture HH26 and HH44 it was the other way round. Expression of SRC-1 was first slightly decreasing from 6 to 24 h and after 72 h the initial expression ratio could be determined.

Nuclear receptor subfamily 1, group I, member 2 (NR1I2, PXR) belongs to the nuclear receptor superfamily which are transcription factors characterized by a ligand-binding domain and a DNA-binding domain. The encoded protein is a transcriptional regulator of the cytochrome P450 gene CYP3A4, binding to the response element of the CYP3A4 promoter as a heterodimer with the 9-cis retinoic acid receptor RXR. It is activated by a range of compounds that induce CYP3A4, including Rifampicin and Dexamethasone. PXR the master regulator of detoxification showed a very similar down-regulation for all three patients. From a starting expression ratio of about one, the expression ratio between Rifampicin und DMSO treated cells was declining to 0.6 in all

cultures meaning a 1.6-fold down-regulation. The down-regulation continued at least till 72 h after induction.

Retinoid X receptor, alpha (RXR α) is one of a whole family of nuclear receptors that mediate the biological effects of retinoids by involvement in retinoic acid-mediated gene activation. These receptors exert their action by binding, as homodimers or heterodimers, to specific sequences in the promoters of target genes like CYP3A4 and regulating their transcription. RXR α was at the beginning higher expressed in culture HH27 (1.5-fold) compared with HH27 (0.9) and HH44 (0.9). The down-regulation after 24 h was observed for all three samples. While the down-regulation for HH27 further increased till 72 hours after induction, the expression ratio of RXR α increased slightly for culture HH26 and HH44.

Nuclear receptor subfamily 0, group B, member 2 (NR0B2, SHP) is an unusual orphan receptor that contains a putative ligand-binding domain but lacks a conventional DNA-binding domain. The receptor is a member of the nuclear hormone receptor family, a group of transcription factors regulated by small hydrophobic hormones, a subset of which do not have known ligands and are referred to as orphan nuclear receptors. The protein has been shown to interact with retinoid and thyroid hormone receptors, inhibiting their ligand-dependent transcriptional activation. In addition, interaction with estrogen receptors has been demonstrated, leading to inhibition of function. Studies suggest that the protein represses nuclear hormone receptor-mediated transactivation via two separate steps: competition with co-activators and the direct effects of its transcriptional repressor function. This interplay with co-activators like NCOA 1 or direct interaction with the nuclear receptors might interfere and down-regulate CYP3A4 expression. SHP which is able to inhibit CYP3A4 expression showed a different behaviour for all three samples. While in culture HH27 it was highly over-expressed after 6 h (2.6-fold) the expression in samples HH26 (0.9) and HH44 (1.2) showed almost no differentially expression. After 24 h of induction the expression ratio of SHP was decreasing in culture HH27 while in the two other cultures (HH26 and HH44) it was increasing reaching for all cultures the same level of about 1.25-fold over-expression. After 72 h the expression of the CYP3A4-expression inhibitor was increasing for samples from culture HH26 (1.4-fold) while it was 2-fold decreasing for samples from HH27 and HH44.

A detailed summary of the observations in the cell cultures from different patients concerning CYP3A4 expression and regulation is provided by Table 52, Figure 102 and Figure 103. Table 52 gives the individual, average and median expression values of genes involved in regulation of CYP3A4. Figure 102 visualizes the detected signals on the arrays as scatter-plots for each patient, treatment and measurement point. From this figure the fold-change and expression change of each gene can be seen. Finally the expression ratios between induced and control cells are plotted over the time to show the time dependent expression change. This dynamic view of the regulatory system is provided in Figure 103.

Table 52: Expression ratios of genes involved in regulation of CYP3A4 expression in hepatocyte cells from three different donors are given. The expression ratio between Rifampicin induced and DMSO treated control cells was measured after 6, 24 and 72 h. The expression level for the following proteins was monitored in three cell cultures (HH26, 27, 44): PPARGC1A, HNF4A, NCOA1, NR1I2, RXRA, CYP3A4 and NR0B2.

PPARGC1A	6	24	72	HNF4A	6	24	72
HH26	1.28	0.67	1.26	HH26	0.91	0.64	0.67
HH27	0.67	1.06	1.04	HH27	2.39	0.39	0.30
HH44	0.99	0.78	1.04	HH44	1.13	0.50	0.57
average	0.98	0.84	1.12	average	1.48	0.51	0.51
median	1.02	0.76	1.17	median	1.23	0.55	0.55

NCOA1	6	24	72	NR1I2	6	24	72
HH26	1.14	0.97	1.07	HH26	0.81	0.68	0.65
HH27	1.04	1.11	0.81	HH27	1.11	0.70	0.63
HH44	0.86	0.83	0.93	HH44	0.97	0.63	0.58
average	1.02	0.97	0.93	average	0.97	0.67	0.62
median	1.04	0.97	0.98	median	0.97	0.68	0.63

RXRA	6	24	72	CYP3A4	6	24	72
HH26	0.89	0.50	0.57	HH26	1.42	2.97	3.94
HH27	1.49	0.69	0.41	HH27	2.45	4.04	1.68
HH44	0.89	0.43	0.63	HH44	1.29	1.94	2.02
average	1.09	0.54	0.53	average	1.72	2.98	2.55
median	0.89	0.50	0.57	median	1.42	2.97	2.02

NR0B2	6	24	72
HH26	0.86	1.21	1.39
HH27	2.64	1.23	0.37
HH44	1.19	1.37	0.80
average	1.56	1.27	0.85
median	1.19	1.23	0.80

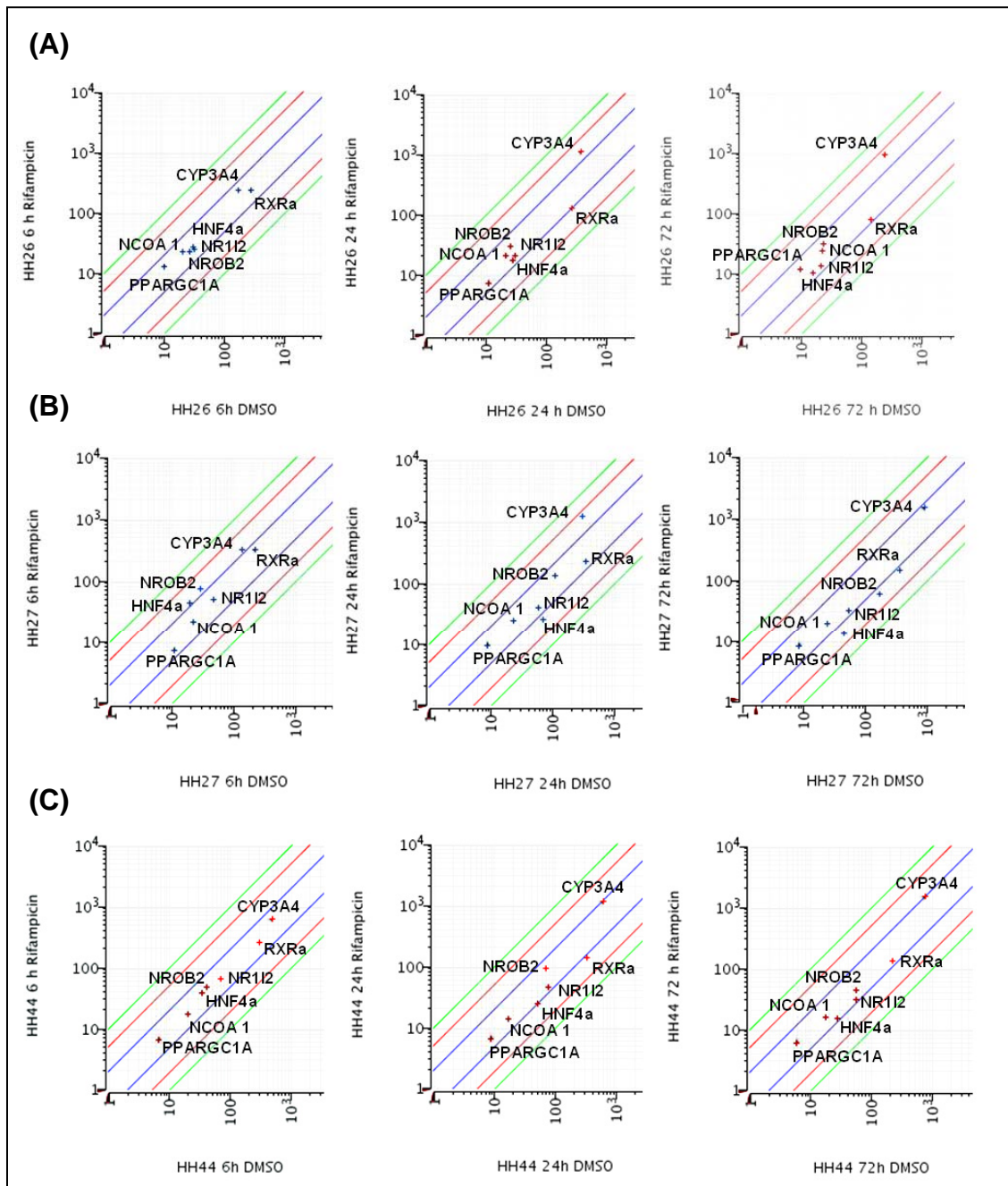


Figure 102: Scatter plot of genes involved in regulation of CYP3A4 expression. The expression levels are given for the Rifampicin treated (Y-axis) and control cells (X-axis). (A) shows the expression after 6, 24 and 72 hours for cell culture HH26, (B) and (C) for the respective points of time for culture HH27 and HH44. The fold change of expression is visualized by colored lines (blue = 2-fold, red = 5-fold and green 10-fold).

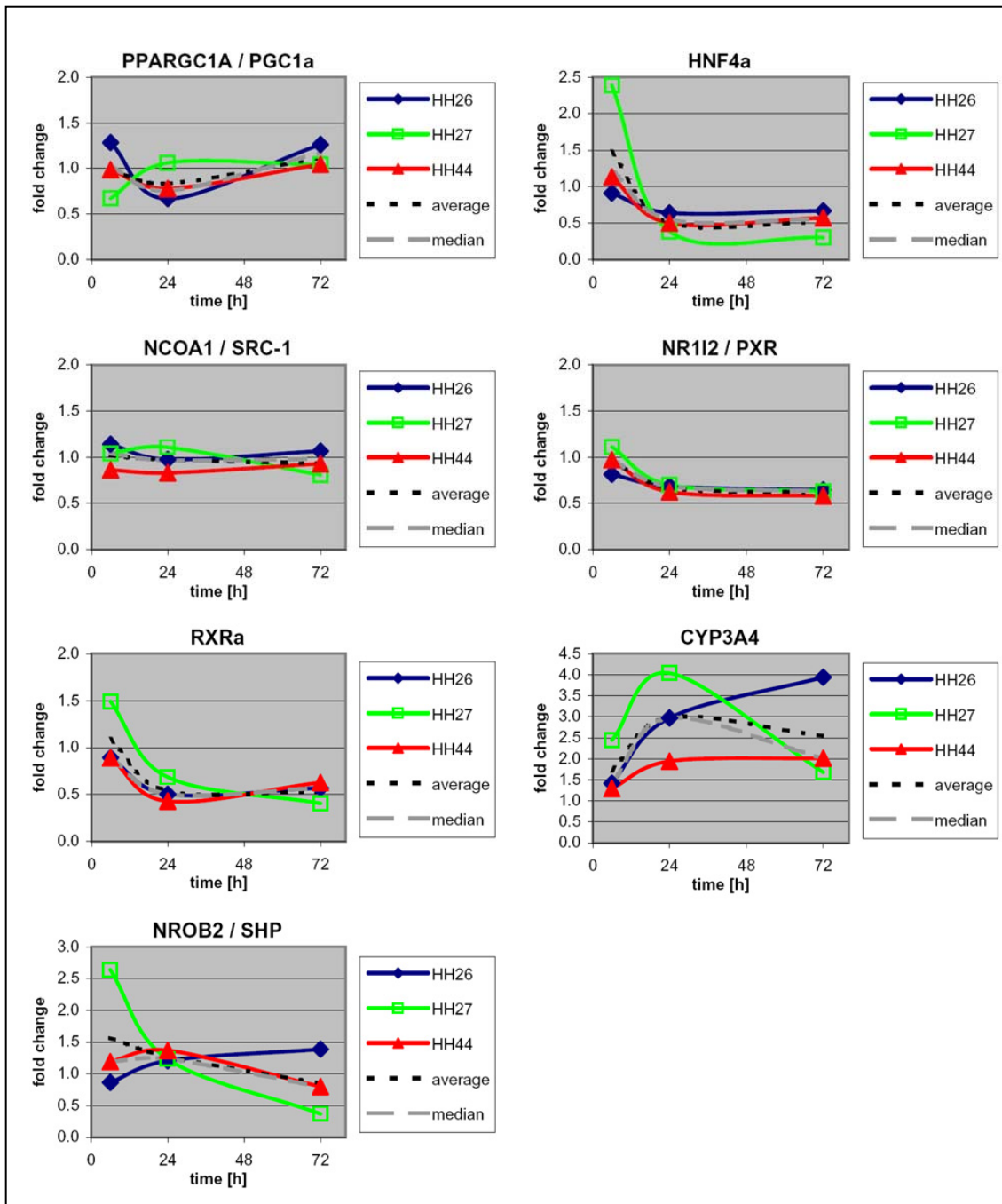


Figure 103: Expression profiles of genes involved in regulation of CYP3A4 expression. The change of expression is given over the time for the three individual donors. For each analysed gene the average and median expression was determined and plotted in the graph as well.

3.8.4 Interaction of Xenobiotic with Endogenous Metabolism Based on Nuclear Receptor Interaction.

In recent studies interactions between drug metabolism and endogenous reactions were revealed. These interactions are based on two reasons. On one hand the endogenous reactions are catalyzed by the same enzymes which are induced or repressed by the drug. Or on the other hand the drug alters the expression or activity of the nuclear receptor which not only regulates the xenobiotic metabolism but is also involved in other regulatory circuits. A good example for this observation is the bile acid metabolism. Here both effects can be found, as nuclear receptor PXR is involved in regulation of bile acid metabolism and CYPs are catalyzing main reaction pathways.

3.8.4.1 Bile Acid Metabolism

As xenobiotic- and bile acid metabolism are both regulated by the nuclear pregnane X receptor the impact of Rifampicin induction on the expression of genes involved in bile acid metabolism was investigated.

The role of nuclear receptors, especially nuclear pregnane X receptor (PXR; NR1I2) and farnesol X receptor in altering the expression levels of genes metabolising bile acids are summarized in Figure 104.

As shown in Figure 104 expression of PXR is declining in induced versus control cells. However the protein seems to be somehow activated because the effects of induction of PXR, a down-regulation of CYP7A1 genes and an up-regulation of CYP3A4 were detected. For culture HH26 after 24 h and HH44 after 72 h the expression of CYP7A1 was detected to be high, compared to all other samples. In the mentioned samples a higher expression was observed in DMSO cell cultures compared to Rifampicin induced cultures. As long as CYP7A1 is high expressed Rifampicin induction seems to down-regulate CYP7A1 expression.

The induction of CYP3A4 was weak after 6 h, however clearly detectable after 24 and 72 h. In culture HH27 and HH44 a weaker but obvious induction occurred in the DMSO treated samples as well.

The expression of FXR was first slightly increasing. After 72 h of induction the expression ratio was again slightly below the initial expression level. PXR was slightly down-regulated after 24 and 72 h of induction.

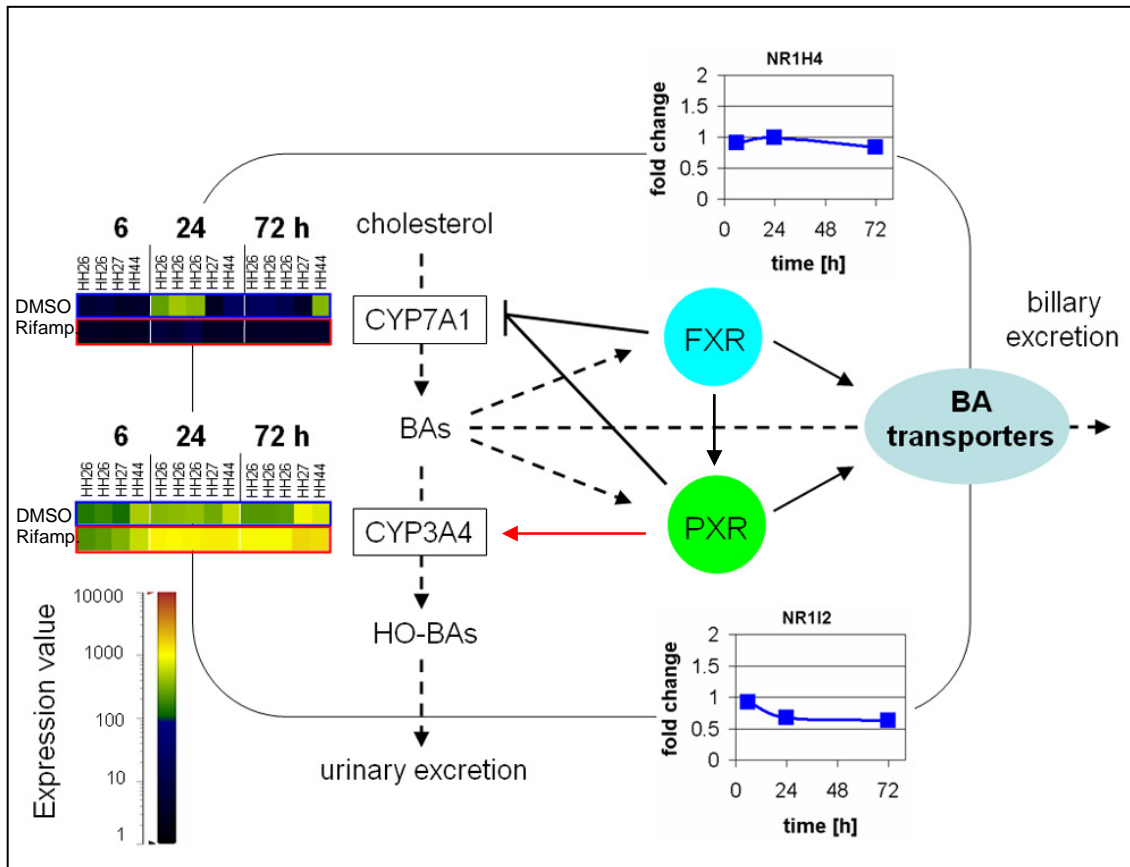


Figure 104: Interplay of bile acid receptors (PXR and FXR). Bile acids (BAs) bind in to PXR and FXR in the liver, resulting in a suppression of CYP7A1 expression. Bile acid synthesis, stimulation of bile acid transporter expression and excretion of bile acids in the bile are reduced. PXR stimulates CYP3A4 expression (red arrow). High CYP3A4 activity leads to formation of hydroxylated bile acids which are excreted from the body into urine. The actions of bile acid on the nuclear receptors and the metabolic pathways are indicated by dotted lines. The expression levels of CYP7A1 and CYP3A4 are given colour coded as tile plots. Each rectangle represents the expression value of the gene in the respective culture mentioned. The upper values are from the DMSO cultures (blue rectangle) and the lower are from the Rifampicin treated culture (red rectangle). For both nuclear receptors the ratio between treated and control samples are plotted as a graph.

A more detailed overview of the expression change of so far known involved genes is given in Figure 105. In part (A) of Figure 105 an overview of the bile acid metabolism and transport is given. Part (B) shows the corresponding time course expression profiles measured with Affymetrix full genome arrays in three cell cultures. A down-regulation of CYP7A1 was measured after 24 h induction with Rifampicin. After 6 h CYP7A1

showed a 1.6-fold (0.65) down-regulation which increased to 14-fold (0.07) after 24 hours and finally reached a 6.5-fold (0.15) down-regulation.

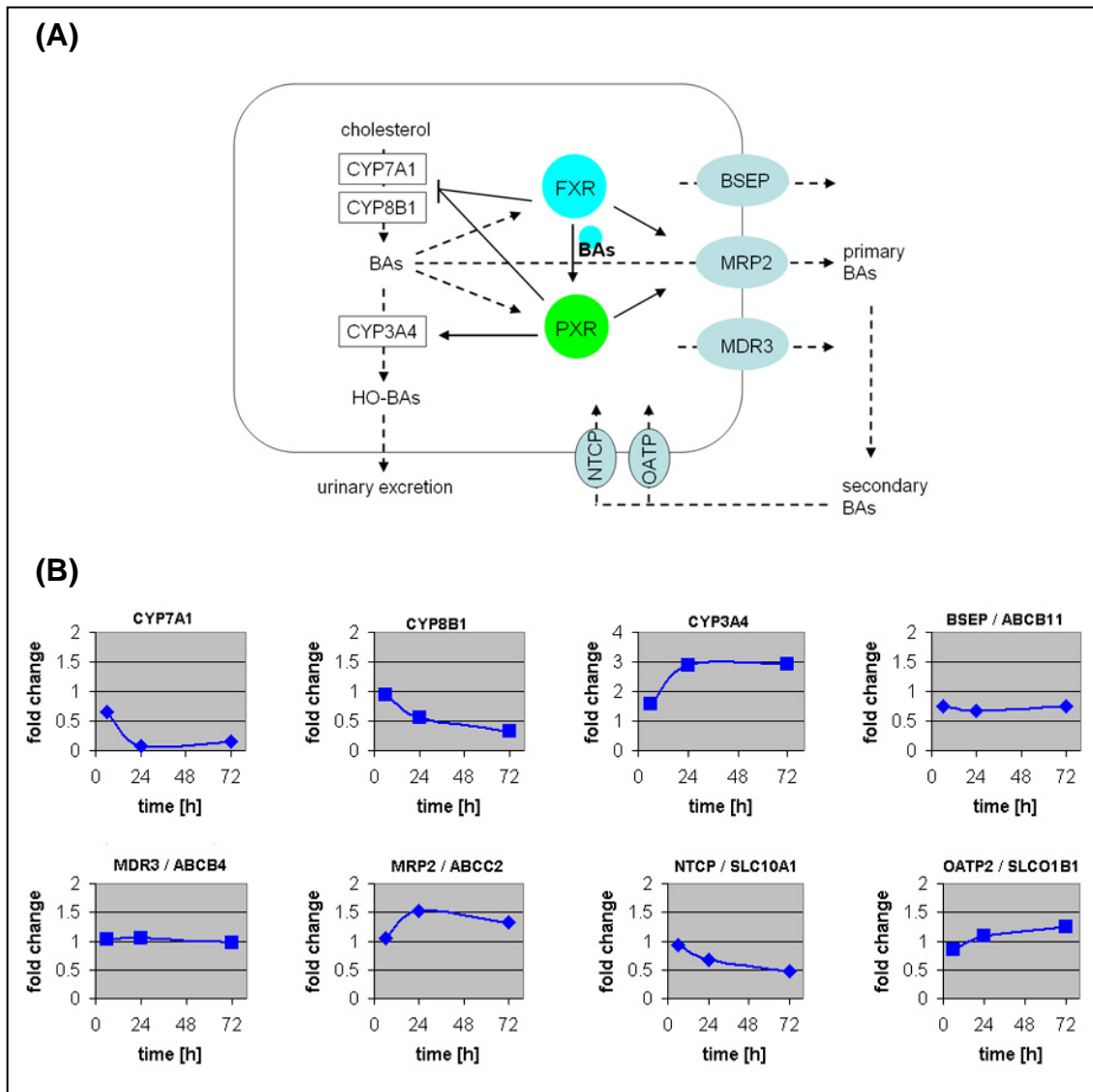


Figure 105: Expression of genes involved in bile acid (BA) formation and circulation. (A) Schematic overview of genes involved in bile acid metabolism. Primary bile acids are synthesized from cholesterol by mainly CYP7A1 and CYP8B1. BAs are detoxified by CYP3A4 into metabolites and transported by bile-salt-excretory pump (BSEP, ABCB11), multidrug resistance protein 3 (MDR3, ABCB4) and multidrug resistance associated protein 2 (MRP2, ABCC2). Primary bile acids in the intestine are converted into secondary bile acids and taken up by hepatocytes through Na⁺-taurocholate cotransporting polypeptide (NTCP, SLC10A1) and Na⁺-independent organic anion transporting polypeptide 2 (OATP2, SLCO1B1). (B) Expression profiles of genes involved in bile acid metabolism. The ratio of gene expression between Rifampicin induced and not induced control cells is given for 6, 24 and 72 h.

Vice versa an increase was detected for CYP3A4. This member of the cytochrome P450 family increased its expression in Rifampicin treated samples from 1.6-fold after

6 h to 2.9-fold for 24 and 72 h samples. CYP8B1 (3-fold) and NTCP (SLC10A1; 2.1-fold) were down-regulated while BSEP (ABCB11) and MDR3 (ABCB4) showed almost no differential expression. The expression of MRP2 (ABBC2; 1.5-fold after 24 h) and OATP2 (SLCO1B1; 1.3-fold after 72 h) was up-regulated. MRP2 was immediately up-regulated after 24 h while OATP2 showed a slowly increase which reached its maximum after 72 h.

3.8.4.2 Interaction of Drug Metabolism with Fatty Acid Degradation

The Rifampicin based induction of genes involved in xenobiotic metabolism has also impact on the regulatory network of lipid metabolism. The following example shows a partially down-regulation of the adipocytokine signalling pathway / PPAR signalling pathway which was not very significant but interesting to notice (Figure 106):

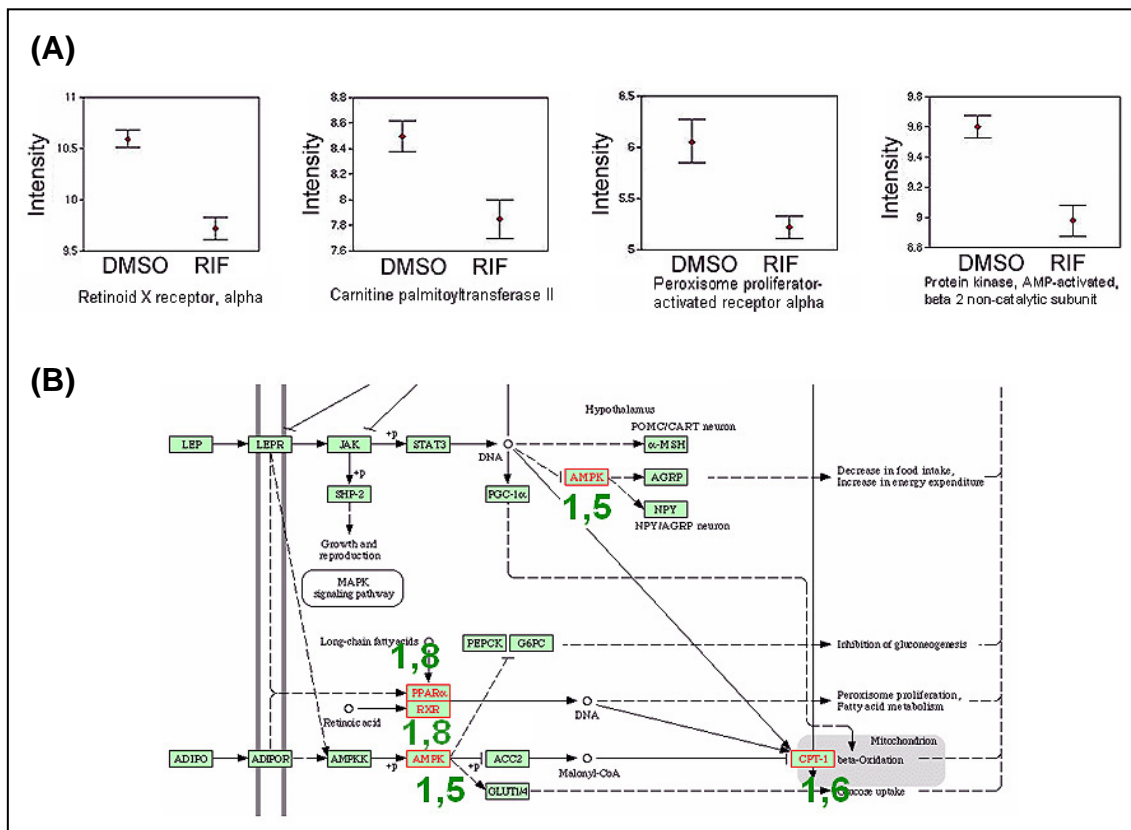


Figure 106: Down-regulation of parts of the adipocytokine signalling pathway in primary human hepatocytes after 24 h treatment with Rifampicin compared to control cells. (A) Expression intensities for RXR α , CPT-1, PPAR α and AMPK detected with DNA microarrays for control (DMSO) and induced (RIF) samples. **(B)** Part of the KEGG pathway showing the regulated genes (red) and their factor of down-regulation (green).

In this experiment retinoid X receptor alpha (RXRa) and peroxisome proliferator-activated receptor alpha (PPARa) showed a 1.8 fold down-regulation after 24 h of Rifampicin induction. Carnitine palmitoyltransferase II (CPT-1) was down-regulated 1.6-fold and Protein kinase, AMP-activated, beta 2 non-catalytic subunit was down-regulated 1.5-fold. These observations clearly demonstrate the interaction between endogenous- and xenobiotic metabolism using similar signalling pathways. Down- or up-regulation of regulatory proteins like nuclear receptors by xenobiotic drugs affect not only the detoxifying enzymes themselves but also common networks of energy metabolism or lipid homeostasis.

3.8.5 Time-Series Approach with Sub-Genome Arrays

In a first "bottom-up" approach the same three time-points (6, 24, 72 h) as for the full genome analyses were investigated using sub-genome microarrays. Samples from three different patients (HH43, HH44 and HH48) were monitored. A summary of the detected differentially expressed genes for each point in time is given in Table 53 - Table 55.

Table 53: Quantitative significant expressed genes in three human hepatocyte cell cultures after 6 h treatment with Rifampicin. Given is the ratio between induced and control measurement for each culture as well as an average ratio. Furthermore the standard deviation of the three biological replicates and the log-ratio (ln) are given.

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	ln-ratio
1	CYP2B6	13.15	1.67	2.36	5.73	6.44	1.75
2	CYP2C	7.54	1.29	1.85	3.56	3.46	1.27
3	ABCC1	0.28	9.32	1.04	3.55	5.01	1.27
4	CYP1A2	7.37	0.50	1.21	3.03	3.78	1.11
5	IGFBP3	1.14	1.93	1.72	1.59	0.41	0.47
6	CYP4A11	2.73	0.57	1.04	1.45	1.14	0.37
7	CYP1B1	1.29	0.78	1.58	1.21	0.41	0.19
8	FOS	0.14	1.41	1.83	1.13	0.88	0.12
9	ABCC5	0.00	0.10	3.27	1.12	1.86	0.12
10	CYP4F3	0.59	1.02	1.53	1.05	0.47	0.05
11	CEBPA	1.83	0.49	0.82	1.05	0.70	0.05
12	FMO5	1.24	0.64	0.93	0.94	0.30	-0.06
13	BAX	1.50	0.72	0.41	0.88	0.56	-0.13
14	PPARG	0.51	0.84	0.98	0.78	0.24	-0.25
15	CCND3	0.93	0.37	0.97	0.76	0.33	-0.28

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
16	STE	1.01	0.52	0.71	0.74	0.25	-0.29
17	ABCC3	1.02	0.66	0.50	0.73	0.27	-0.32
18	SLC22A7	0.51	0.87	0.67	0.68	0.18	-0.38
19	IL15	0.96	0.64	0.43	0.68	0.27	-0.39
20	HKG YWHAZ	0.40	0.97	0.31	0.56	0.36	-0.58
21	CYP2E1	0.09	0.66	0.91	0.55	0.42	-0.59
22	TEP1	0.33	0.42	0.83	0.53	0.27	-0.64
23	IL6R	0.73	0.58	0.27	0.53	0.23	-0.64

More information on the mentioned genes in this section like gene name, Genbank and Swissprot accession number as well as general biological function can be referred from Table 64 in the appendix.

The experiments performed with DualChip microarrays revealed a high variation between samples from different donors. Cell culture HH48 showed a different expression profile with very strong inducibility compared to HH43 and HH44. Despite all differences some common patterns could be observed for all samples. Up-regulated in all three points in time for all three patients were CYP2B6 and CYP2C. CYP2E1 was down-regulated in all patients for all points in time. For the late measurement points (24 and 72 hours) GSA, CYP1A1, EPHX1 and UGT1A1 showed a higher expression. NR1I3, IGF1 and CDKN1B were similarly down-regulated.

Table 54: Quantitative significant expressed genes in three human hepatocyte cell cultures after 24 h treatment with Rifampicin. Given is the ratio between induced and control measurement for each culture as well as an average ratio. Furthermore the standard deviation of the three biological replicates and the log-ratio (ln) are given.

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
1	CYP2A	1.34	0.27	138.48	46.70	79.49	3.84
2	GSA	1.21	2.00	44.18	15.80	24.58	2.76
3	CYP2C	1.86	1.92	28.48	10.75	15.35	2.38
4	CYP3A	2.47	1.27	13.30	5.68	6.63	1.74
5	EHHADH	0.37	0.27	13.94	4.86	7.87	1.58
6	CYP2B6	5.70	2.41	6.14	4.75	2.04	1.56
7	EPHX1	1.85	1.63	6.74	3.41	2.89	1.23
8	IL1A	0.14	1.43	8.42	3.33	4.46	1.20
9	SLC21A6	0.63	0.81	8.25	3.23	4.35	1.17
10	UGT1A1	1.45	1.61	6.44	3.17	2.84	1.15
11	MAPK10	1.81	5.05	1.73	2.86	1.90	1.05
12	ABCC6	-	4.38	0.79	2.59	2.54	0.95
13	CEBPA	0.64	0.63	6.46	2.57	3.37	0.95
14	CYP1A1	3.38	1.69	2.57	2.55	0.84	0.93

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
15	ABCC4	0.51	1.55	5.39	2.48	2.57	0.91
16	SLC22A1	0.63	0.43	6.35	2.47	3.36	0.90
17	SLC22A2	3.23	0.00	4.05	2.43	2.14	0.89
18	IL11RA	4.93	0.57	1.78	2.42	2.25	0.89
19	SULT1A	0.48	0.92	5.62	2.34	2.85	0.85
20	CCNE1	0.44	2.10	4.18	2.24	1.87	0.81
21	SLC22A7	0.19	0.25	6.22	2.22	3.46	0.80
22	GSTT2	0.28	0.89	4.87	2.01	2.49	0.70
23	MAPK8	3.20	1.96	0.68	1.95	1.26	0.67
24	NR1I2	0.30	0.28	5.09	1.89	2.77	0.64
25	ABCB1	1.81	1.89	1.89	1.86	0.04	0.62
26	CYP2D6	0.52	0.58	4.44	1.85	2.24	0.61
27	ABCG2	1.37	0.82	3.28	1.82	1.30	0.60
28	CCND1	0.54	0.88	4.03	1.82	1.93	0.60
29	CYP1A2	0.47	0.62	4.20	1.76	2.11	0.57
30	GSTM5	-	1.54	1.87	1.70	0.23	0.53
31	ABCB11	1.47	0.06	3.34	1.62	1.65	0.48
32	TGFBR II	0.47	0.86	3.37	1.57	1.58	0.45
33	GSTT1	0.65	0.65	3.10	1.47	1.41	0.38
34	PTGS2	-	2.85	0.08	1.46	1.96	0.38
35	CYP4A11	0.54	0.31	3.49	1.45	1.77	0.37
36	GSM	0.42	0.84	2.89	1.38	1.32	0.33
37	TNFRSF7	1.98	0.80	1.30	1.36	0.59	0.31
38	UGT2B7	0.58	0.66	2.83	1.36	1.28	0.30
39	FMO5	0.55	0.46	2.95	1.32	1.41	0.28
40	CYP4F3	0.65	0.52	2.77	1.31	1.27	0.27
41	HSPA4	0.66	0.89	2.37	1.30	0.93	0.27
42	NAT 1/2	0.53	0.61	2.76	1.30	1.27	0.26
43	TP53	0.33	1.60	1.96	1.30	0.86	0.26
44	IGF1R	0.15	0.68	2.99	1.28	1.51	0.24
45	CYP1B1	1.87	1.27	0.64	1.26	0.61	0.23
46	SLC15A1	0.62	0.64	2.42	1.23	1.03	0.20
47	FMO4	0.43	0.62	2.60	1.22	1.20	0.19
48	DDIT3	0.63	0.74	2.26	1.21	0.91	0.19
49	EPHX2	0.47	0.47	2.54	1.16	1.20	0.15
50	HSPB1	0.94	0.51	2.02	1.16	0.78	0.14
51	MYCBP	0.51	0.96	1.88	1.12	0.70	0.11
52	MAPK9	0.42	1.32	1.61	1.12	0.62	0.11
53	MGMT	0.59	0.87	1.74	1.07	0.60	0.07
54	ABCC1	0.65	1.73	0.66	1.01	0.62	0.01
55	COL6A2	0.05	0.22	2.73	1.00	1.50	0.00
56	CCND3	0.41	0.87	1.71	1.00	0.66	0.00
57	GSTM3	0.55	1.12	1.29	0.99	0.39	-0.01
58	FOS	1.58	1.30	0.03	0.97	0.83	-0.03
59	MAOA	0.60	0.94	1.18	0.91	0.29	-0.10
60	PPARG	0.34	0.54	1.66	0.85	0.71	-0.17
61	HGFAC	0.30	0.51	1.68	0.83	0.74	-0.19
62	IL6R	0.61	0.32	1.55	0.83	0.64	-0.19

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
63	COMT	0.66	0.65	1.10	0.80	0.26	-0.22
64	CASP8	0.70	0.63	1.04	0.79	0.22	-0.23
65	LOC51064	0.55	0.60	1.11	0.75	0.31	-0.28
66	MET	0.44	1.03	0.75	0.74	0.29	-0.30
67	FN1	1.07	0.61	0.48	0.72	0.31	-0.33
68	GRLF1	0.53	0.14	1.35	0.67	0.62	-0.40
69	CYP4B1	0.50	0.56	0.89	0.65	0.21	-0.43
70	PPARD	0.00	0.54	1.28	0.60	0.64	-0.50
71	CDKN2A	0.00	0.23	1.57	0.60	0.84	-0.51
72	CDKN1B	0.57	0.47	0.57	0.54	0.06	-0.62
73	HMOX1	0.52	0.48	0.49	0.50	0.02	-0.70
74	BAX	0.64	0.65	0.16	0.48	0.28	-0.73
75	SLC21A9	0.39	0.44	-	0.42	0.04	-0.87
76	IGF1	0.24	0.35	0.65	0.41	0.21	-0.88
77	CCND2	0.07	0.13	0.95	0.38	0.49	-0.96
78	NR1I3	0.25	0.21	0.67	0.38	0.25	-0.98
79	IGFBP5	0.12	0.66	0.20	0.33	0.29	-1.12
80	CYP2E1	0.19	0.09	0.31	0.20	0.11	-1.62
81	STE	0.07	0.07	-	0.07	0.00	-2.62

Table 55: Quantitative significant expressed genes in three human hepatocyte cell cultures after 72 h treatment with Rifampicin. Given is the ratio between induced and control measurement for each culture as well as an average ratio. Furthermore the standard deviation of the three biological replicates and the log-ratio (ln) are given.

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
1	CYP2B6	13.09	2.52	4.97	6.86	5.53	1.93
2	CCNB1	18.71	1.11	0.67	6.83	10.29	1.92
3	CCND2	9.55	1.64	2.31	4.50	4.39	1.50
4	GSA	9.19	2.04	1.52	4.25	4.28	1.45
5	CYP2C	7.09	1.92	2.93	3.98	2.74	1.38
6	CYP1A2	8.33	0.61	0.90	3.28	4.38	1.19
7	ABCB1	5.41	1.96	1.95	3.11	2.00	1.13
8	SLC21A8	7.29	0.38	0.44	2.70	3.97	0.99
9	CYP1A1	3.58	1.03	1.87	2.16	1.30	0.77
10	EPHX1	2.79	1.70	1.82	2.10	0.60	0.74
11	CCNE1	4.96	0.54	0.80	2.10	2.48	0.74
12	UGT1A1	2.69	1.64	1.29	1.87	0.73	0.63
13	CYP2A	3.14	0.28	1.26	1.56	1.46	0.44
14	GSTP1	1.57	0.65	1.43	1.21	0.50	0.19
15	ABCC4	0.77	1.12	1.66	1.18	0.45	0.17
16	CEBPA	1.65	0.66	1.09	1.13	0.50	0.13
17	TNFRSF7	0.73	0.51	1.86	1.03	0.73	0.03
18	GSTT1	1.45	0.63	0.90	0.99	0.42	-0.01
19	CASP8	1.54	0.62	0.69	0.95	0.51	-0.05
20	CCNA1	1.66	0.53	0.18	0.79	0.77	-0.23

No.	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
21	CYP2D6	1.17	0.58	0.55	0.77	0.35	-0.26
22	CYP4A11	1.63	0.34	0.24	0.74	0.78	-0.31
23	FMO5	1.11	0.44	0.62	0.72	0.34	-0.32
24	SLC22A1	1.23	0.42	0.41	0.69	0.47	-0.38
25	NAT 1/2	0.65	0.65	0.71	0.67	0.03	-0.40
26	IL11RA	0.84	0.48	0.66	0.66	0.18	-0.41
27	LOC51064	0.71	0.58	0.57	0.62	0.08	-0.48
28	CYP4F3	0.61	0.55	0.69	0.61	0.07	-0.49
29	SLC15A1	0.46	0.66	0.69	0.61	0.13	-0.50
30	EPHX2	0.68	0.51	0.61	0.60	0.08	-0.51
31	EHHADH	0.94	0.34	0.45	0.58	0.32	-0.55
32	IL15	0.40	0.58	0.70	0.56	0.15	-0.58
33	ABCC6	0.00	1.13	0.53	0.55	0.56	-0.59
34	FMO4	0.58	0.54	0.53	0.55	0.03	-0.60
35	RARA	0.40	0.54	0.69	0.54	0.15	-0.61
36	GSM	0.49	0.44	0.69	0.54	0.13	-0.62
37	PPARG	0.28	0.44	0.88	0.53	0.31	-0.63
38	IL6R	0.64	0.31	0.64	0.53	0.19	-0.64
39	COL6A2	0.36	0.52	0.69	0.52	0.17	-0.65
40	PPARA	0.23	0.46	0.73	0.48	0.25	-0.74
41	TEP1	0.10	0.60	0.71	0.47	0.32	-0.75
42	CDKN1B	0.09	0.59	0.72	0.47	0.33	-0.76
43	IGF1	0.31	0.35	0.71	0.45	0.22	-0.79
44	BCLX	0.00	0.38	0.97	0.45	0.49	-0.80
45	NR1I3	0.58	0.19	0.28	0.35	0.20	-1.05
46	SLC22A7	0.46	0.22	0.29	0.32	0.12	-1.13
47	CDKN2A	0.34	0.07	0.29	0.23	0.15	-1.45
48	NR1I2	0.17	0.12	-	0.15	0.03	-1.93
49	CYP2E1	0.11	0.11	0.20	0.14	0.05	-1.98

After 6 hours of induction only one gene (SLC22A7) showed a similar expression in all biological replicates. Four genes (HMOX1, ABCB1, CDKN1B and CYP2E1) were similar expressed after 24 hours. 72 hours after induction the expression level of 15 differentially expressed genes was similar (Table 56).

Except ABCB1 which was similarly up-regulated in all samples after 24 hours (1.8 to 1.9) all genes showing a similar expression were down-regulated (Table 56). CYP2E1 showed a very significant down-regulation of 0.1 to 0.3 in all three patient samples for 24 and 72 hours. SLC22A7 was down-regulated after 6 hours and 72 hours of induction with 0.5 to 0.9 after 6 and slightly between 0.2 and 0.5 after 72 hours. The smallest biological variance in all three samples showed HMOX1 which was 0.5 down-

regulated. After 72 h of induction FMO4 and NAT1/2 showed a very similar down-regulation of 0.5 to 0.6 and 0.7.

Table 56: Given are twenty genes showing the smallest variance of gene-expression in cell cultures of three different patients. Point in time, the gene ID and the ratios between induced and control samples are shown. From the ratios of the three patients the average was calculated and the standard deviation was determined, indicating similar expression.

No.	time	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
1	24 h	HMOX1	0.52	0.48	0.49	0.50	0.02	-0.70
2	72 h	FMO4	0.58	0.54	0.53	0.55	0.03	-0.60
3	72 h	NAT 1/2	0.65	0.65	0.71	0.67	0.03	-0.40
4	24 h	ABCB1	1.81	1.89	1.89	1.86	0.04	0.62
5	72 h	CYP2E1	0.11	0.11	0.20	0.14	0.05	-1.98
6	24 h	CDKN1B	0.57	0.47	0.57	0.54	0.06	-0.62
7	72 h	CYP4F3	0.61	0.55	0.69	0.61	0.07	-0.49
8	72 h	EPHX2	0.68	0.51	0.61	0.60	0.08	-0.51
9	24 h	CYP2E1	0.19	0.09	0.31	0.20	0.11	-1.62
10	72 h	SLC22A7	0.46	0.22	0.29	0.32	0.12	-1.13
11	72 h	SLC15A1	0.46	0.66	0.69	0.61	0.13	-0.50
12	72 h	GSM	0.49	0.44	0.69	0.54	0.13	-0.62
13	72 h	CDKN2A	0.34	0.07	0.29	0.23	0.15	-1.45
14	72 h	RARA	0.40	0.54	0.69	0.54	0.15	-0.61
15	72 h	IL15	0.40	0.58	0.70	0.56	0.15	-0.58
16	72 h	COL6A2	0.36	0.52	0.69	0.52	0.17	-0.65
17	6 h	SLC22A7	0.51	0.87	0.67	0.68	0.18	-0.38
18	72 h	IL11RA	0.84	0.48	0.66	0.66	0.18	-0.41
19	72 h	IL6R	0.64	0.31	0.64	0.53	0.19	-0.64
20	72 h	NR1I3	0.58	0.19	0.28	0.35	0.20	-1.05

In contrast to similar expression of down-regulated genes, most of the up-regulated genes showed a very individual change of expression in different patients. Table 57 summarizes the 20 differentially expressed genes showing the highest variance between biological replicates. All genes were induced and therefore stronger expressed in Rifampicin treated cultures compared to control cultures leading to a positive log-ratio. As cell culture 48 showed a very strong inducibility compared to the other cultures, the variance for the CYP genes like CYP2A, CYP2C, CYP3A and CYP2B6 was quite high. But not for all genes the highest expression was observed for culture HH48. For example CYP2B6 which was expressed quite different after 6 and 72 hours had the highest expression in culture HH43 (13.1) while in culture 44 and 48 it was also up-regulated, but only 1.7- to 5.0- fold. CYP2C was up-regulated after 6 and 24 hours with an increasing variance from 3.5 after 6 h of induction to 15.4 after 24 hours. After 6 h

cell culture HH43 showed the highest expression for CYP2C while the highest expression after 24 h was monitored for cell culture HH48. A similar phenomenon was observed for GSA for the 24 and 72 hour measurements. After 24 hours of induction culture 48 had the highest expression (44.18) leading to a standard deviation of 24.6, while after 72 hours the expression of GSA in culture 44 caused the high variance of 4.3.

Table 57: Given are twenty genes showing the highest variance of gene-expression in cell cultures of three different patients. Given is the time, gene ID and the ratios between induced and control samples. From the ratios of the three patients the average was calculated and the standard deviation was determined, indicating similar expression.

No.	Time	GeneID	Ratio 43	Ratio 44	Ratio 48	Average	STABW	In-ratio
1	24 h	CYP2A	1.34	0.27	138.48	46.70	79.49	3.84
2	24 h	GSA	1.21	2.00	44.18	15.80	24.58	2.76
3	24 h	CYP2C	1.86	1.92	28.48	10.75	15.35	2.38
4	72 h	CCNB1	18.71	1.11	0.67	6.83	10.29	1.92
5	24 h	EHHADH	0.37	0.27	13.94	4.86	7.87	1.58
6	24 h	CYP3A	2.47	1.27	13.30	5.68	6.63	1.74
7	6 h	CYP2B6	13.15	1.67	2.36	5.73	6.44	1.75
8	72 h	CYP2B6	13.09	2.52	4.97	6.86	5.53	1.93
9	6 h	ABCC1	0.28	9.32	1.04	3.55	5.01	1.27
10	24 h	IL1A	0.14	1.43	8.42	3.33	4.46	1.20
11	72 h	CCND2	9.55	1.64	2.31	4.50	4.39	1.50
12	72 h	CYP1A2	8.33	0.61	0.90	3.28	4.38	1.19
13	24 h	SLC21A6	0.63	0.81	8.25	3.23	4.35	1.17
14	72 h	GSA	9.19	2.04	1.52	4.25	4.28	1.45
15	72 h	SLC21A8	7.29	0.38	0.44	2.70	3.97	0.99
16	6 h	CYP1A2	7.37	0.50	1.21	3.03	3.78	1.11
17	24 h	SLC22A7	0.19	0.25	6.22	2.22	3.46	0.80
18	6 h	CYP2C	7.54	1.29	1.85	3.56	3.46	1.27
19	24 h	CEBPA	0.64	0.63	6.46	2.57	3.37	0.95
20	24 h	SLC22A1	0.63	0.43	6.35	2.47	3.36	0.90

To get a more detailed insight in the dynamic behaviour of gene expression in human hepatocyte cells, a sub-genome expression profiling with a higher sampling frequency of cell culture 48 was analysed. An overview of the experimental procedure is given in Figure 107.

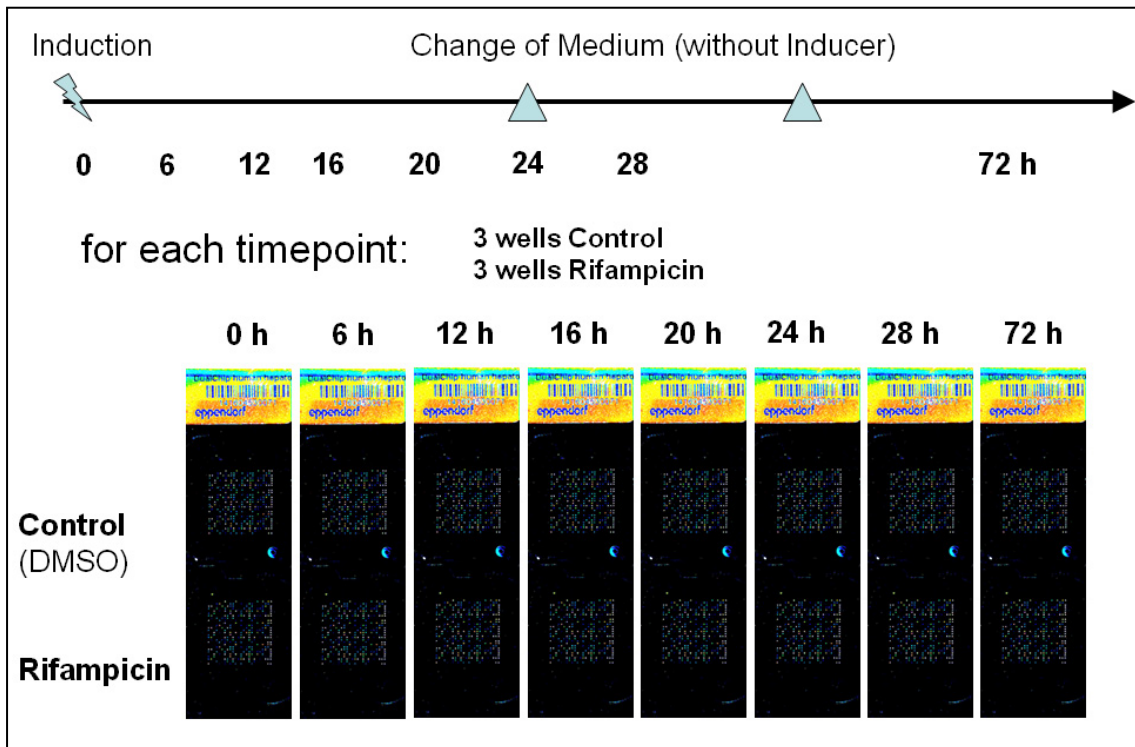


Figure 107: Experimental setup for sub-genome expression profiling using hepatocyte cell cultures and Eppendorf DualChips. The experimental design was based on time matched controls. Each sampling point (6, 12, 16, 20, 24, 28 and 72 h) consisted of induced (Rifampicin) and not induced (DMSO) samples on one slide. Induction with Rifampicin was done at time point zero. Every 24 h medium was removed and changed with medium not containing the inducer Rifampicin (marked by triangles on the timeline). This strategy was followed to see the dynamic response of the transcription machinery.

Samples of induced (Rifampicin) and not induced (DMSO) cells were harvested after 6, 12, 16, 20, 24, 28 and 72 h of induction. With the exchange of medium after 24 hours the inducer (Rifampicin) was removed. This strategy allowed the investigation of induction in detail during the first part of the time series. After removing Rifampicin the response in gene expression on this change was monitored in the late points of measurement.

The detection of mRNA levels was performed on Hepato DualChips from the company Eppendorf allowing the detection of 151 liver and detoxification relevant genes. An overview of the total expression change in cell culture HH48 is given in Figure 108.

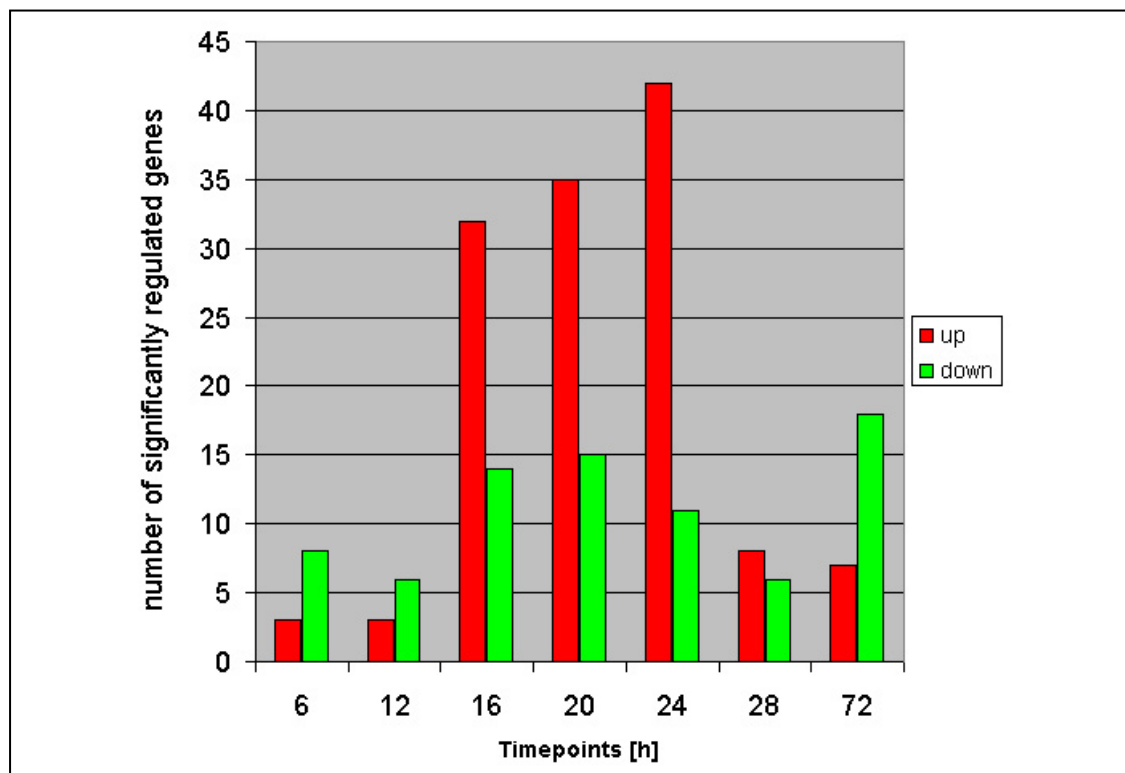


Figure 108: Total number of differentially expressed genes after several time points. After six and twelve hours of induction with Rifampicin only a small number of genes is up- and downregulated in hepatocyte cells. The strongest induction of genes is between 16 and 24 hours. After 24 h the medium was exchanged without Rifampicin. After 28 h, four hours after withdrawing the inducer the expression status is more or less as at the beginning. After 72 h the number of downregulated genes is stronger than that of induced ones.

After 6 and 12 hours of induction only three genes showed a significant up-regulation while eight and six genes were down-regulated. 16 hours after induction 32 genes were up-regulated and 14 genes down-regulated. After 20 hours of induction the number of differentially expressed genes increased to 35 up-regulated and 15 down-regulated genes. The highest observed change in expression occurred after 24 hours of induction. 35% of all genes on the sub-genome arrays were differentially expressed. 42 genes showed an up-regulation while eleven genes were down-regulated. Removal of Rifampicin after 24 hours led to a decreased change in expression levels. 28 hours after induction and 4 hours after exchanging the medium without inducer eight genes were up-regulated and six were down-regulated. After 72 hours of total cultivation time and after induction as well as 48 hours after removing the inducing substance, seven genes were up-regulated and 18 genes detected to be down-regulated.

A detailed summary of the differentially expressed genes for every point in time is given in the following figures (Figure 109 - Figure 115). Figure 109 shows the genes with the highest change in expression level after 6 hours of induction. The regulation score was determined by calculating the ratio between reference and experiment intensities on the array. The ratio is multiplied by the normalization factor obtained from the housekeeping- and internal standard normalization. Finally the ratio is log transformed using the natural logarithm (ln). As an example the calculation for CYP2B6 is given:

Signal intensity reference: 14.648

Signal intensity experiment: 23.078

Calculating the ratio of signal intensities: $ratio = \frac{i_{exp}}{i_{ref}} = \frac{23.078}{14.648} = 1.58$

Multiplication with normalization factor (1.47): $normalized\ ratio = 1.58 \cdot 1.47 = 2.32$

Log-transformation of ratio: $\ln(normalized\ ratio) = "log" = \ln(2.32) = \underline{\underline{0.8}}$

Cytochrome P450 2B6 (CYP2B6) had a log-ratio of 0.8 indicating a 2.3 fold up-regulation while Cytochrome P450 2C (CYP2C) and murine osteosarcoma viral oncogene homolog (FOS) were up-regulated 1.8 times. For Interleukin 6 receptor (IL6R) a ratio of 0.26 was calculated resulting in $\ln(0.26) = -1.3$ -fold change or in other words a 1.3-fold down-regulation.

In part B of Figure 109 the down-regulated genes are given. IL6R (0.26, 1.3-fold), cyclin A (CCNA1, 0.30 fold), aryl hydrocarbon receptor (AHR, 0.35), BCL2-associated X protein (BAX, 0.39), interleukin 15 (IL15, 0.42), multidrug resistance-associated protein 3 (ABCC3, 0.48), N-acetyltransferase 1/2 (NAT1/2, 0.57) and monoamine oxidase B (MAOB, 0.58) were down-regulated after 6 h of induction.

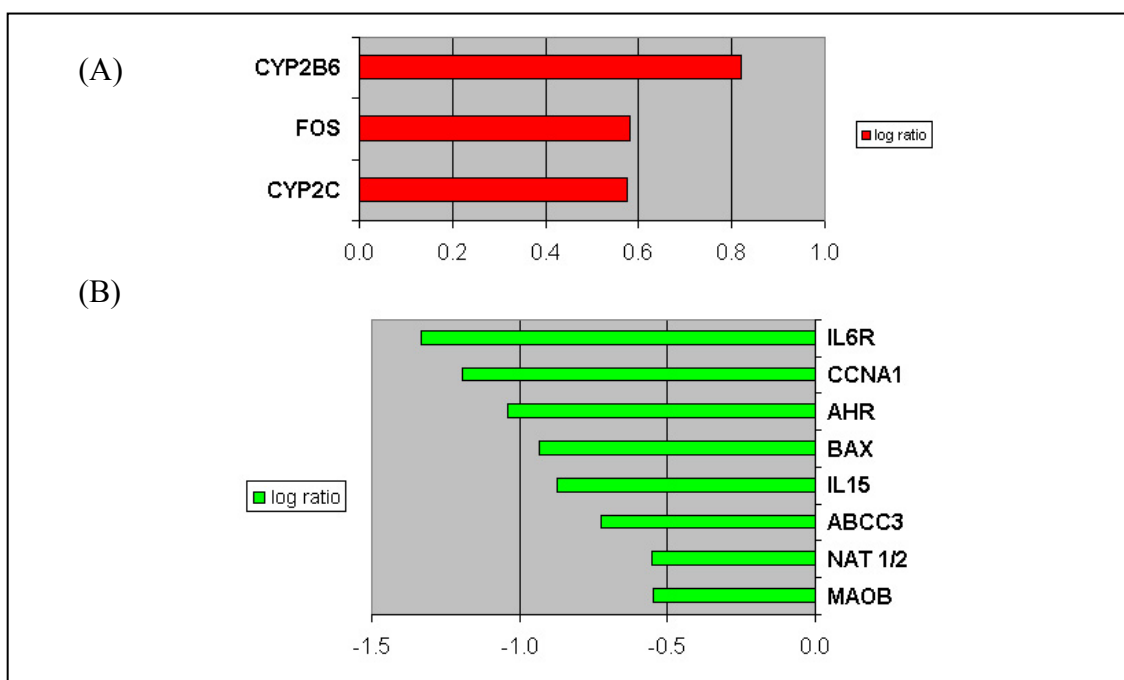


Figure 109: Differentially regulated genes after 6 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval (A) Three genes showing a more than two-fold up-regulation. (B) Eight genes were found to be more than two-fold down-regulated.

After 12 hours of induction CYP2B6 and CYP2C were still up-regulated with 2.5-fold and 1.8-fold as given in Figure 110. In addition multidrug resistance-associated protein 2 (ABCC2) was expressed 2.0-fold higher in induced cells compared to control cells. The previously higher expressed FOS showed no significant differential expression after 12 hours. From the eight down-regulated genes after 6 hours of induction, only CCNA1 was still down-regulated after 12 hours with a ratio of 0.5. In addition glutathione S-transferase M3 (GSTM3, 0.24), ATP-binding cassette, sub-family B (MDR/TAP, ABCB11, 0.40), organic anion transporter B (SLC21A9, 0.5) and Cytochrome P450 4A11 (CYP4A11, 0.5) were down-regulated.

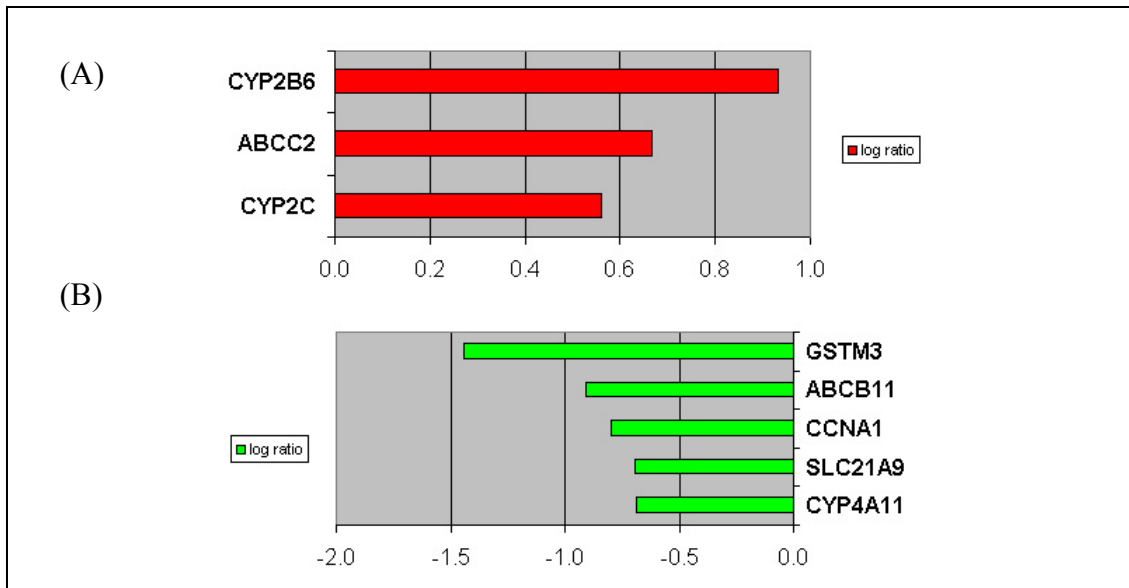


Figure 110: Differentially regulated genes after 12 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval (A) Three genes showing a more than two-fold up-regulation. (B) Eight genes were found to be more than two-fold down-regulated.

After 16 hours of induction the number of differentially expressed genes increased significantly. This increase in expression level was measured independently three times. The variance of the three measurements is indicated as error bars in Figure 111. As previously monitored CYP2C and CYP2B6 were more than two-fold up-regulated (18.9- and 6.8-fold). CYP2A had an even higher expression with a fold change of 4.0 corresponding to a log ratio of almost 53. CYP3A and CYP4A11 were also up-regulated 9.8 (2.3-fold) and 6.9 (1.9-fold) times. Furthermore the following genes showed a more than two-fold increase in expression when comparing induced versus control samples: organic anion transporter B (SLC21A9, 9.3), epoxide hydrolase microsomal (EPHX1, 9.2), glutathione S-transferase A2 (GSA, 8.0), CCAAT/enhancer-binding protein (CEBPA, 7.6), Tumour protein p53 (TP53, 7.3), solute carrier family 22 (organic anion transporter), member 7 (SLC22A7, 6.9), hydratase/3-hydroxyacyl coenzyme A dehydrogenase (EHHADH, 4.9), sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1 (SULT2A1, 4.9), O-6-methylguanine-DNA methyltransferase (MGMT, 4.6), apolipoprotein E (APOE, 4.4), cyclin-dependent kinase 4 (CDK4, 4.4), UDP glucuronosyltransferase 1 family,

polypeptide A (UGT1A, 3.9), glutathione transferase zeta 1 (GSTZ1, 3.8), sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 (SULT1A, 3.8) and organic cation transporter I (SLC22A1, 3.5).

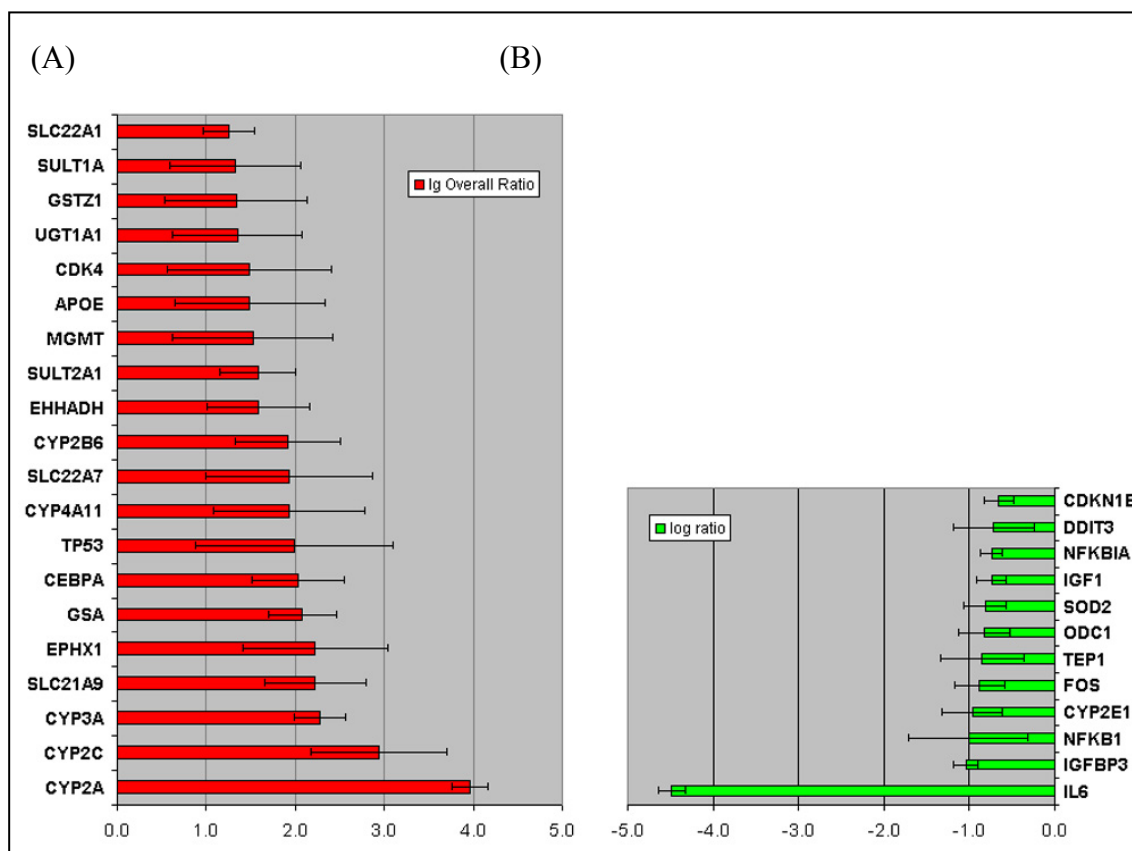


Figure 111: Differentially regulated genes after 16 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval (A) 20 genes showing a larger than two-fold up-regulation. (B) Twelve genes were found to be more than two-fold down-regulated.

A very strong down-regulation with a high significance between the three observations was detected for IL6 with a log ratio of over four (4-fold down-regulation) correlating to a decrease of 0.01 in expression. In addition eleven more genes were down-regulated after 16 h of induction with Rifampicin: insulin growth factor binding protein 3 (IGFBP3, 0.4), nuclear factor of kappa B (NFKB1, 0.4), Cytochrome P450 2E1 (CYP2E1, 0.4), FOS (0.4), Telomerase associated protein 1 (TEP1, 0.4), Ornithine decarboxylase1 (ODC1, 0.4), superoxide dismutase 2 (SOD2, 0.4), Insulin like growth factor1 (IGF1, 0.5), nuclear factor of kappa light polypeptide gene enhancer in B-cells

inhibitor, alpha (NFKBIA, 0.5), DNA-damage-inducible transcript 3 (DDIT3, 0.5), cyclin dependent kinase inhibitor 1B (CDKN1B, 0.5).

Figure 112 gives an overview of the most dominant up- (A) and down-regulated genes (B) after 20 hours of induction.

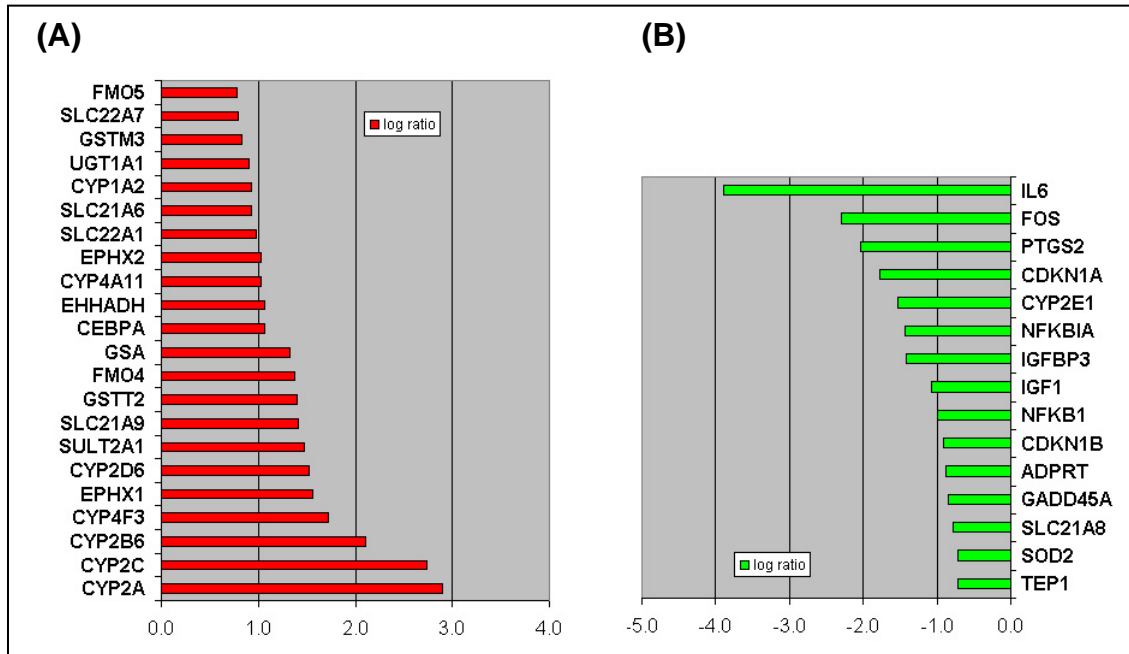


Figure 112: Differentially regulated genes after 20 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval (A) 22 genes showing a more than two-fold up-regulation. (B) 15 genes were found to be more than two-fold down-regulated.

Especially the members of the cytochrome P450 family are strongly induced. CYP2A, 2C, 2B6 had a ratio of 18, 15 and 8 indicating fold change (ln-ratio) larger than two between induced and control sample. CYP2C and 2B6 kept the induction status measured for the 12 h and 16 h measurement points in this cell culture. CYP4F3 was 5.6-fold increased. Epoxide hydrolase microsomal (EPHX1), sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1 (SULT2A1) and glutathione S-transferase theta 2 (GSTT2) showed had expression ratios of 4.8 (1.6-fold), 4.3 (1.5-fold), and 4.0 (1.4-fold) and where again shown to be up-regulated after 20 hours as well as in the previous point in time. For the first time up-regulated in this experiment were CYP2D6 1.5 fold (4.6) and organic anion transporter B (SLC21A9, 1.4 fold, 4.1). A 1.4-fold induction for FMO4 (4.0) and 1.3-fold for GSA (3.8) was

monitored. Further genes detected with an increased expression after 20 h were EHHADH (2.9) and CYP4A11 (2.8) which were also detected already after 16 hours of induction, epoxide hydrolase soluble (EPHX2, 2.8), SLC22A1 (2.7; also after 16 hours up-regulated), SLC21A6 (2.5), cytochrome P450 1A2 (CYP1A2, 2.5), UGT1A1 (2.5; also after 16 hours up-regulated), glutathione S-transferase M3 (GSTM3, 2.3), SLC22A7 (2.2; also after 16 hours up-regulated), flavin containing monooxygenase 5 (FMO5, 2.2).

The strongest down-regulation after 20 hours was again observed for IL6 with a factor of 0.02 (3.9-fold), and FOS with a ratio of 0.1 (2.3-fold). Both genes were also strongly down-regulated after 16 hours of induction. After 20 hours additionally prostaglandin endoperoxidase synthase 2 (PTGS2) showed a reduced expression by 0.1. Further genes measured with a decrease in expression were cyclin dependent kinase inhibitor 1A CDKN1A (0.2) and the previously detected CYP2E1 (0.2), nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (NFKBIA, 0.2), insulin-like growth factor binding protein 3 (IGFBP3, 0.2), Insulin like growth factor1 (IGF1, 0.3), Nuclear factor of kappa B (NFKB1, 0.4), Cyclin dependent kinase inhibitor 1B (CDKN1B, 0.4), superoxide dismutase 2, mitochondrial (SOD2, 0.5) and telomerase associated protein 1 (TEP1, 0.5). For the first time down-regulated in this experiment after 20 hours of induction were poly (ADP-ribose) polymerase family, member 1 (ADPRT, 0.4), growth arrest and DNA-damage-inducible, alpha (GADD45A, 0.4), solute carrier organic anion transporter family, member 1B3 (SLCO1B3 previously SLC21A8, 0.46).

For the 24 hour measurement point after induction three replicates were measured to determine the variance of the system. The variance is plotted in Figure 113 as error bar for each differentially expressed gene.

The samples obtained after 24 hours of induction showed the largest number of genes altering their expression. 42 genes showed an expression ratio larger than two while 11 genes were down-regulated with a ratio larger 0.5.

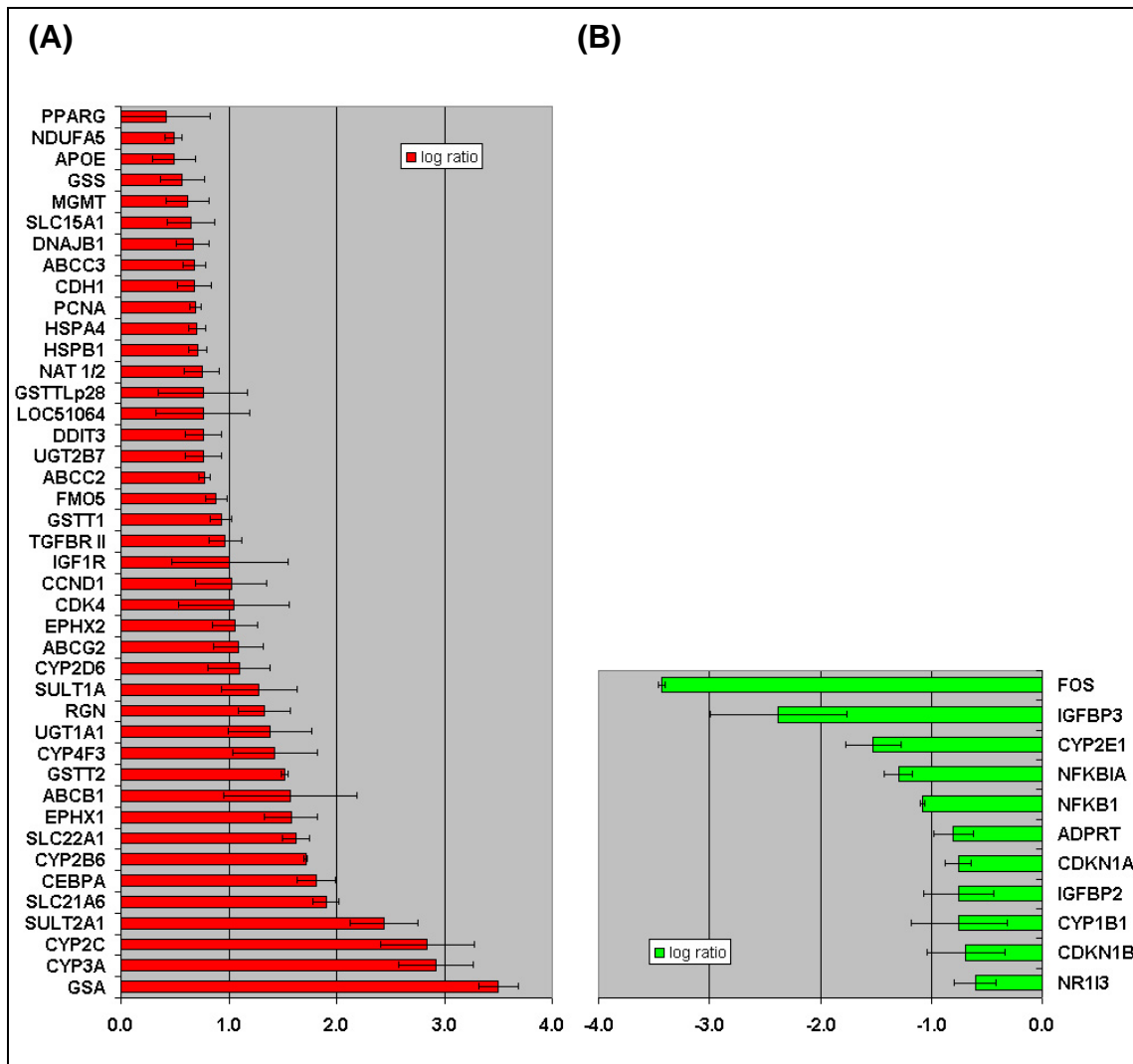


Figure 113: Differentially regulated genes after 24 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval. The given values are an average from six measurements of three technical replicates/arrays with spotted probes in triplicates. The variance between these measurements is indicated by error bars. (A) 42 genes showing a larger than two-fold up-regulation. (B) 11 genes were found to be larger than two-fold down-regulated.

The highest up-regulation after 24 hours showed glutathione S-transferase A2 (GSA,) CYP3A, CYP2C and sulfotransferase family, cytosolic, 2A, dehydroepiandrosterone (DHEA)-preferring, member 1 (SULT2A1) with a fold-change larger two. Organic anion transporter C (SLC21A6), CCAAT/enhancer-binding protein (CEBPA), CYP2B6, organic cation transporter I (SLC22A1), epoxide hydrolase microsomal (EPHX1), ATP-binding cassette, sub-family B (MDR/TAP), member 1 (ABCB1) and GSTT2 showed a higher than 1.5-fold increase in expression level (ratio of 4.5).

Additional genes having an increased expression were UDP glucuronosyltransferase 1 family, polypeptide A (UGT1A), regucalcin (senescence marker protein-30, RGN), sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 (SULT1A), cytochrome P450 2D6 (CYP2D6), ATP-binding cassette, sub-family G (WHITE), member 2 (ABCG2), epoxide hydrolase soluble (EPHX2), cyclin-dependent kinase 4 (CDK4), cyclin D1 (CCND1), insulin like growth factor1 receptor (IGF1R), transforming growth factor beta II (TGFB β II), glutathione S-transferase theta 1 (GSTT1), flavin containing monooxygenase 5 (FMO5), Multidrug resistance-associated protein 2 (ABCC2), UDP glucuronosyltransferase 2 family, polypeptide B7 (UGT2B), DNA-damage-inducible transcript 3 (DDIT3), glutathione S-transferase kappa 1 (LOC51064, GSTK1), glutathione S-transferase omega 1 (GSTO1, GSTTLp28), N-acetyltransferase 1/2 (NAT 1/2), heat shock 27kD protein1 (HSPB1), Heat shock 40kD protein1 (DNAJB1), heat shock 70kD protein1 (HSPA4), proliferating cell nuclear antigen (PCNA), cadherin 12 (CDH1), multidrug resistance-associated protein 3 (ABCC3), solute carrier family 15 (oligopeptide transporter), member 1 (SLC15A1), O-6-methylguanine-DNA methyltransferase (MGMT), Glutathione synthetase (GSS), apolipoprotein E (APOE), NADPH oxidoreductase (NDUFA5) and peroxisome proliferator-activated receptor gamma (PPARG).

The eleven down-regulated genes were v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS), insulin growth factor binding protein 3 (IGFBP3) (as in the previous 20 h measurement point), cytochrome P450 2E1 (CYP2E1), nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (NFKBIA), nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) which were shown to be down-regulated in the previous point in time. Furthermore poly (ADP-ribose) polymerase family, member 1 (ADPRT), cyclin dependent kinase inhibitor 1A (CDKN1A), insulin growth factor binding protein2 (IGFBP2), cytochrome P450 1B1 (CYP1B1), cyclin dependent kinase inhibitor 1B (CDKN1B) and Constitutive androstane receptor (NR1I3, CAR) were down-regulated.

Withdrawing the medium containing Rifampicin and adding medium without inducer after 24 h led to decrease in expression change. Altogether only 13 genes showed a change in their expression four hours after this perturbation. The genes affected with a ratio of at least 0.5 are given in Figure 114.

As in all previous points in time CYP2B6 and CYP2C were up-regulated (1.9-fold, ratio 6.7 and 1.1-fold, ratio 2.9). Cytochrome P450 1A1 (CYP1A1) so far not differentially expressed was up-regulated 1.9-fold (ratio 6.4) after 28 h induction and 4 hours after removal of the inducer. Epoxide hydrolase microsomal (EPHX1) and sulfotransferase family, cytosolic, 2A, member 1 (SULT2A1) showing an up-regulation after 20 and 24 hours were also higher expressed after 28 hours. ATP-binding cassette, sub-family B (ABCB1) which was also up-regulated after 24 h of induction showed still a 0.9-fold increase in expression after 28 hours. Multidrug resistance-associated protein 2 (ABCC2) was also higher expressed in the induced cells compared to the control cells.

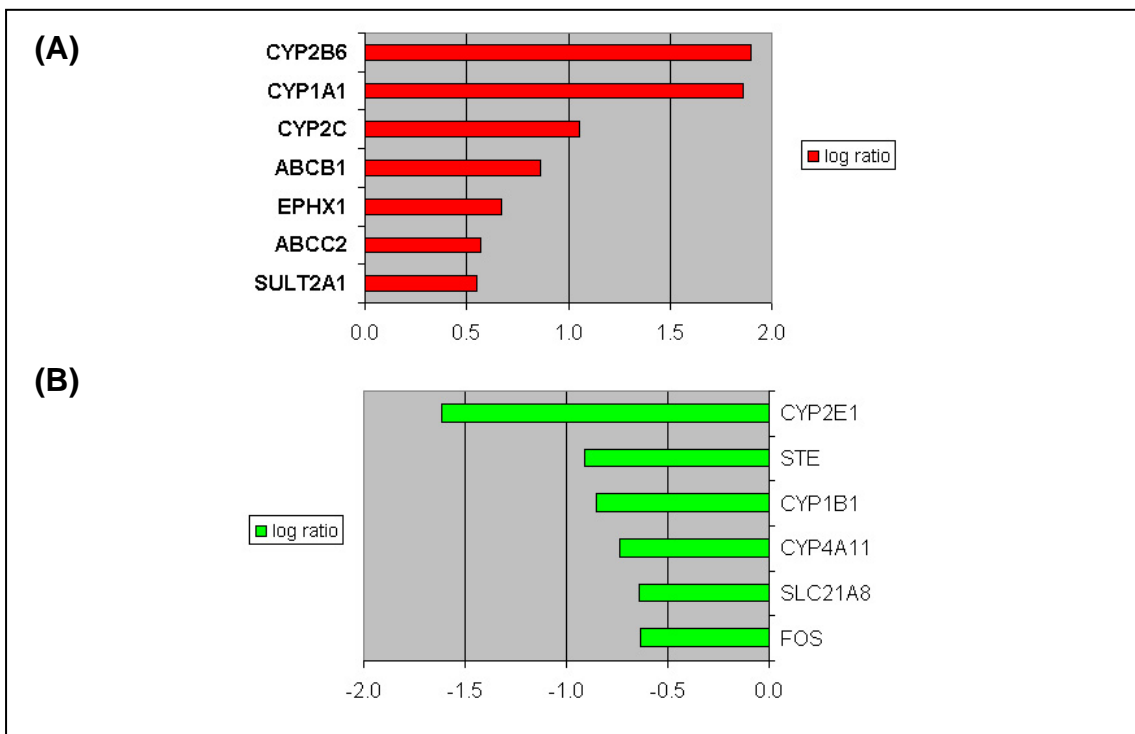


Figure 114: Differentially regulated genes after 28 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. The medium was changed after 24 h without Rifampicin. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval. (A) Seven genes showing a larger than two-fold up-regulation. (B) Six genes were found to be more than two-fold down-regulated.

The most constant down-regulated gene in this cell-culture was v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS) which was monitored to be down-regulated after 20, 24 and 28 hours. However after 28 hours in culture and 4 hours after

removal of the inducer Rifampicin only a 0.5 fold decrease was monitored. CYP2E1 and CYP1B1 were down-regulated as in the previous measurement point. Organic anion transporter 8 (SLC21A8) showed also a reduced expression (0.5 fold) as after 20 hours of induction. In addition sulfotransferase family 1E, estrogen-preferring, member 1 (STE, SULT1E1) previously not measured as differentially expressed showed a 0.4 fold down-regulation after 28 hours.

After 72 hours of induction the number of down-regulated genes exceeded the number of up-regulated genes. A summary of the expression changes is given in Figure 115. Five genes showed an up-regulation including CYP2B6 (1.5-fold, ratio 4.7) and CYP2C (1.0-fold, ratio 2.7) which were up-regulated in all previous time-points as well and CYP1A1 (0.5-fold, ratio 1.7) which was detected to be up-regulated in the previous point in time. In addition tumour necrosis factor receptor (TNFRSF7) and cyclin D2 (CCND2) showed a ratio of 1.8 and 2.5 corresponding to 0.6 and 0.9-fold induction.

The genes sulfotransferase family 1E, estrogen-preferring, member 1 (STE, SULT1E1, 0.02), Cytochrome P450 2E1 (CYP2E1, 0.2), Constitutive androstane receptor (NR1H3, CAR, 0.3) and Organic anion transporter 8 (SLC21A8, 0.4) were detected as down-regulated after 72 h as in a previous point in time. Several more genes showed a down-regulation after 72 h of induction and 48 h after withdrawing Rifampicin as well. These were hepatocyte nuclear factor 4, alpha (HNF4a, 0.02), glutathione S-transferase theta 2 (GSTT2, 0.1), cyclin A (CCNA1, 0.2; which was previously shown to be down-regulated after 12 h), Cytochrome P450 4A11 (CYP4A11, 0.2), Cyclin dependent kinase inhibitor 2A (CDKN2A, 0.3), solute carrier family 22 (organic anion transporter), member 7 (SLC22A7, OAT2, 0.3), enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase (EHHADH, 0.4; which was also up-regulated after 20 h), flavin containing monooxygenase 4 (FMO4, 0.5), multidrug resistance-associated protein 6 (ABCC6, 0.5), Cytochrome P450 2D6 (CYP2D6, 0.5), glutathione S-transferase kappa 1, LOC51064, GST13, 0.5; up-regulated after 24 h), epoxide hydrolase soluble (EPHX2) and flavin containing monooxygenase 5 (FMO5, both 0.6; up-regulated also after 20 and 24 hours), regucalcin senescence marker protein-30 (RGN, 0.6) and NADPH oxidoreductase (NDUFA5, 0.6).

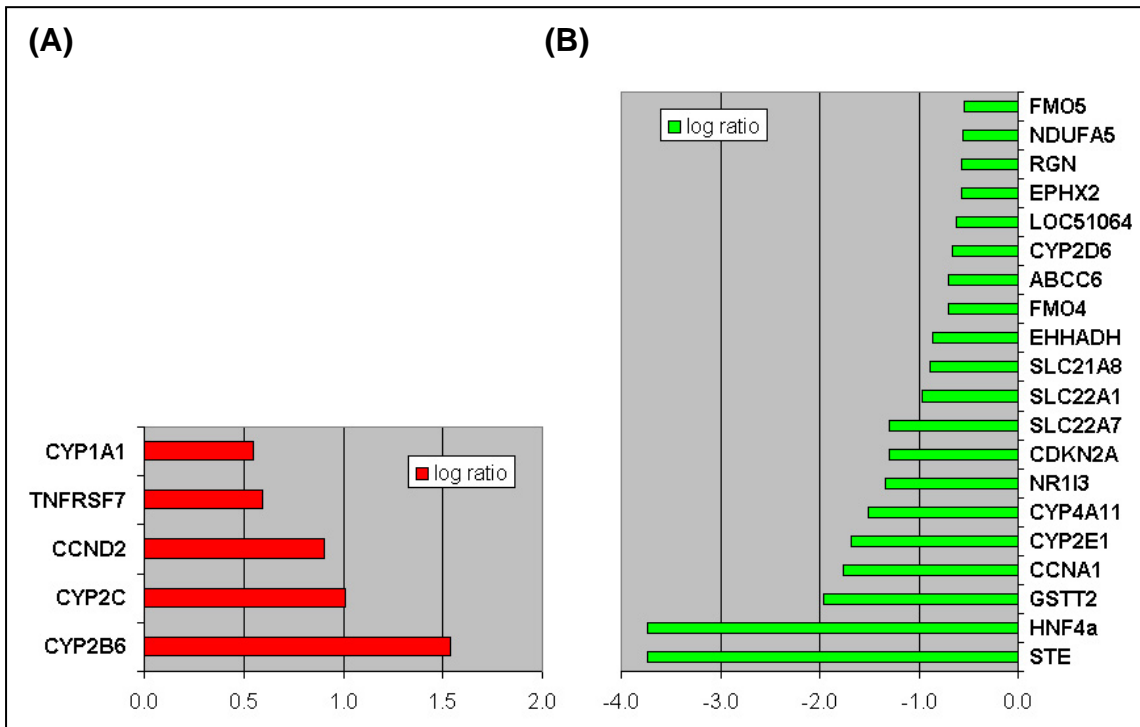


Figure 115: Differentially regulated genes after 72 h Rifampicin induction in primary human hepatocytes (cell culture HH48) analysed using Eppendorf DualChips. The medium was changed after 24 and 48 h without Rifampicin. Genes mentioned here show the highest change at this time point. The ratios given here are outside the 0.99 confidence interval. (A) Five genes showing a higher than two-fold up-regulation. (B) 20 genes were found to be more than two-fold down-regulated.

The sub-genome measurements with seven analysed points in time allowed the first dynamic analysis of the response of gene expression to a perturbation with Rifampicin. As an example the expression level of three CYPs induced by Rifampicin is given in Figure 116. An obvious induction of 2- to 2.9-fold (ratio 8 - 19) is visible after 16 hours of cultivation of primary cells with Rifampicin. The induction is maintained as long as the inducer is present in the medium. Four hours after removing the inducer (Rifampicin), the expression levels of CYP3A and CYP2C are going down, while CYP2B6 expression remains similar for at least 4 hours after removal of inducer.

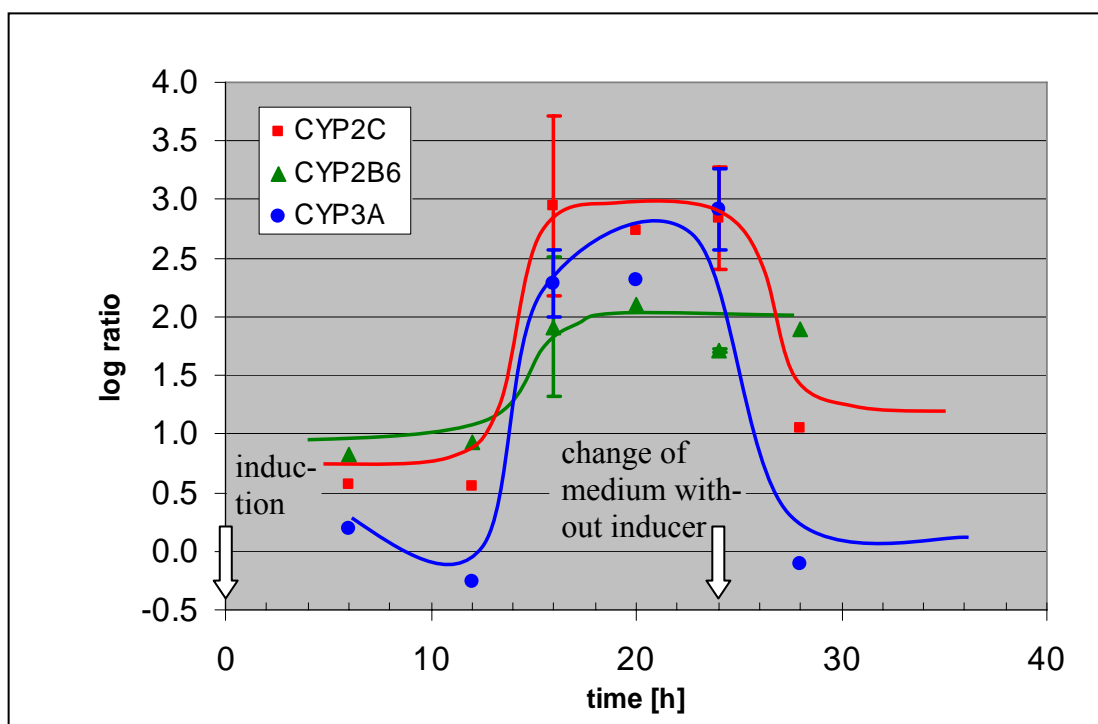


Figure 116: Dynamic expression of members of the Cytochrome P450 family (CYPs) measured with sub-genome microarrays from Eppendorf. The induction was performed at time-point zero and medium was changed after 24 hours without inducer. Samples were analysed after 6, 12, 16, 20, 24 and 28 hours. After 16 hours an induction for all three shown CYPs (2C, 2B6 and 3A) was measured. Four hours after removing the inducer (28 hour time point). CYP3A (●) showed a basal expression level while the expression of CYP2C (■) was about half to the fully induced state and CYP2B6 (▲) was still fully induced.

To even enhance this dynamic view of expression change due to induction, a real-time PCR based assay was performed to allow an even higher sampling frequency.

3.8.6 Time-Series Approach Based on Real-Time PCR

To measure an even closer sampling rate, a real-time PCR strategy was established. The relative expression levels of CYP3A4 and PXR were determined for an induced versus a control cell culture (HH29) for 0.5, 1, 4, 8, 12, 16, 24, 30, 36, 42 and 48 hours after induction.

In parallel CYP3A4 protein concentration was determined by western-blotting for 6, 12, 24 and 72 h. The results of this analysis are shown in Figure 117. After adding medium with inducer (Rifampicin, marked with arrows in Figure 117) at 0 and 24 hours an inducing effect, resulting in a higher expression of the respective genes (CYP3A4 and

PXR) was monitored. Both genes showed a higher expression in Rifampicin treated cultures compared to reference cultures. While CYP3A4 was strongly induced PXR expression was only slightly affected and showed only a weak increase of expression level. CYP3A4 mRNA ratio reached its maximum after 30 h. The increase of CYP3A4 mRNA was also reflected by CYP3A4 protein concentration. Highest amount of CYP3A4 protein was detected after 48 h.

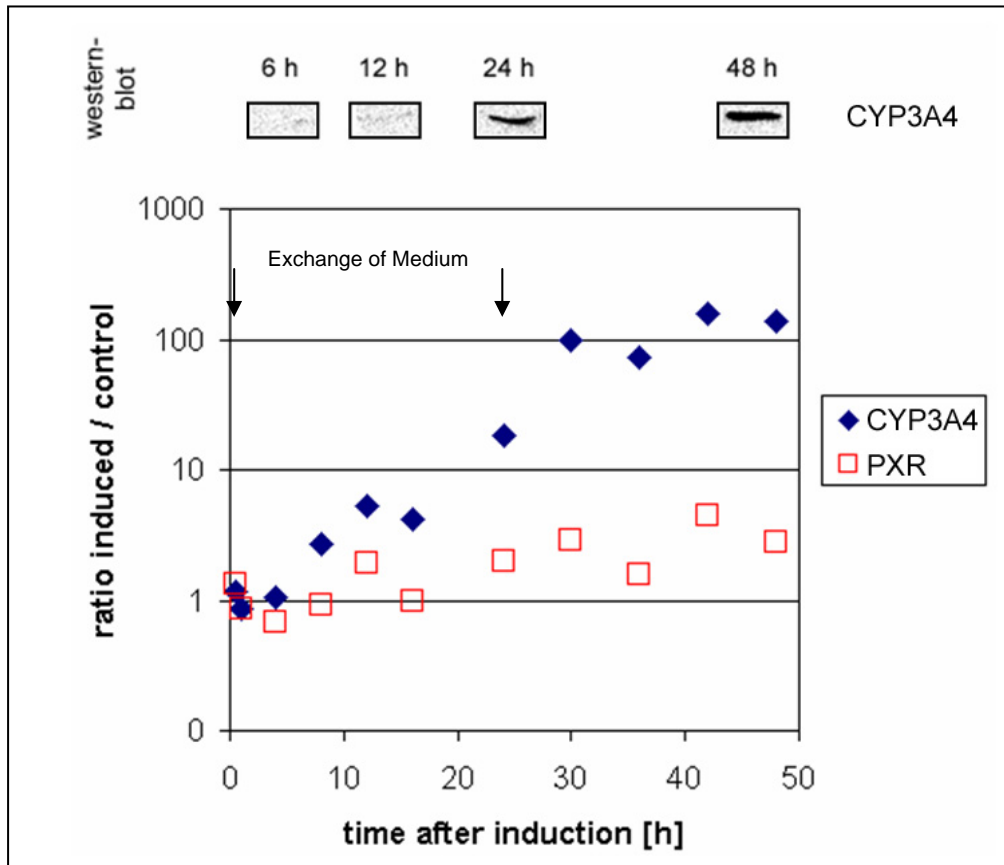


Figure 117: Time series induction experiment monitoring the expression of CYP3A4 and PXR after stimulating human hepatocyte cell cultures (HH29) with Rifampicin. In the upper part of the graph the increasing amount of CYP3A4 protein is shown detected by western-blot. The lower graph displays the mRNA levels of CYP3A4 (◆) and PXR (□) determined by real-time PCR. Arrows mark exchange of medium with Rifampicin for experiment samples and without Rifampicin, only containing DMSO for reference samples.

4 Discussion

Microarray analysis is a very powerful tool which can provide information about the expression of thousands of genes in parallel. As it is common in systems biology to globally analyze the behaviour of a cell the method of choice are experiments with full genome DNA microarrays. The obtained holistic data sets on expression levels of all genes allow generation of new biological knowledge and broaden the view on specific areas of cellular functions. An alternative way to increase knowledge is by analysing specific cellular functions in more detail to gain new insights in already known mechanisms. Both microarray based strategies were used in the work presented here. As microarray technology is a complex technique a lot of work is necessary to establish the application of gene expression microarrays in the laboratory. Many technical parameters and methods, from sample preparation to data analysis need to be optimized, to minimize technical errors associated with this kind of experiments.

Despite the difficulties especially in data evaluation and annotation, DNA-microarrays are the only high throughput method allowing the analysis of the cellular response on mRNA level for all human genes in parallel. This method therefore delivers an indispensable part for data generation in systems biology approaches.

4.1 Methodical considerations

To obtain best possible results, experimental procedures have to be optimized and standardized. This includes development of standard operating protocols (SOPs) for all experimental steps from cell culture, induction experiments, RNA extraction, amplification or reverse transcription and labeling as well as slide handling and quantification. At the end the most time consuming and important part of a microarray study the data interpretation needs also a certain degree of standardization to obtain reproducible and meaningful results.

4.1.1 Obtaining Human Samples for Microarray Gene Expression Profiling

It was a quite difficult task to work with human samples. On one hand the availability of human samples is a limiting factor and on the other hand the individual variability is extremely high in the human population. One of the results from this study is that it is very important for studies planning to work with human cell cultures to assure the necessary amount of cell material. Especially when working with human samples which in contrast to e.g. microorganisms show a high variability between different samples. Furthermore it is not possible to have a not perturbed control like it is possible when working with mice, rats or microorganisms. Special care has to be taken when analysing data from expression profiles obtained with human samples, as all donors suffered from illness and where somehow treated already. This pre-treatment of the patients including with their medical history and the stress for the cells being removed from the organ liver has a strong effect on the gene expression pattern. Nevertheless the idea behind our analysis was by using time match controls to minimize these side effects caused by previous induction or repression of genes. For each measurement point a not induced control sample was analysed next to the induced one to see the effects of induction and not of any illness or pre-treatment. This is a possible way to deal with the problem of not having a healthy, unaltered control. However it still has to be considered that the inducibility and starting conditions of each sample were different. Another disadvantage of this method is that the amount of needed cell material is higher compared working with one reference sample for all points in time. Despite all challenges and difficulties the pattern of induction caused by Rifampicin was generally consistent among different hepatocyte cultures. However especially the analysis with the low-density Eppendorf arrays revealed that the absolute induction or repression varied considerably. Differences in liver condition, preparation and transport caused a large variability mainly in the induction but also in the repression of genes between hepatocyte preparations. This effects were already observed and discussed in many publications (Silva *et al.*, 1998; LeCluyse *et al.*, 1999; LeCluyse *et al.*, 2000; Meunier *et al.*, 2000)

4.1.2 RNA Isolation, Amplification and Labeling

One of the most important factors that determine quality of gene expression experiments is purity and integrity of RNA samples. Since RNA quality mainly relies on the handling during extraction, four preparation methods were compared. Three kit based extractions and phenol chloroform purification of nucleic acids were used. Based on time, cost, handling, reproducibility and yield the kit from Qiagen was considered to be the best. However the two other kits (Agilent and MacheryNagel) as well as the phenol chloroform based method delivered satisfying results of pure, not degraded RNA as well.

Cultivation of primary cells growing in a monolayer allowed isolation of only small amounts of total RNA (generally several micrograms for each sample), thus making an RNA amplification step prior to microarray analysis necessary. The amplification combined with indirect labeling had several advantages. The amount of needed RNA for a successful microarray experiment could be obtained from 1 μ g of starting material. Furthermore no limitation of reverse transcription reaction based on incorporation of large fluorescent dyes (Wiegant *et al.*, 1996) was given as the amplification protocol implies indirect labeling with biotin modified nucleotides. Also dye bias could be excluded as similar biotinylated nucleotides were used for both samples to be compared and hybridized spatial separated on one slide or on two different arrays. The disadvantage of this method is that based on the same labeling the samples are not hybridized on the same array. However this disadvantage is more important for self spotted arrays than for commercial ones which are normally more reproducible. Additionally more time and more money are needed for the experiments. Nevertheless to obtain the best results with small starting RNA amounts we followed this strategy of sample amplification.

The amplification of small amounts of RNA especially obtained from microdissections is quite common and widely accepted (Upson *et al.*, 2004, Luzzi *et al.*, 2005, McClain *et al.*, 2005, Schindler *et al.*, 2005). Presently, the most common amplification methods are linear amplification by *in-vitro* transcription (IVT) and PCR based exponential amplification with a "switching mechanism at the 5' end of the RNA transcript" (SMART). Both methods were tested for reliability and suitability. In this work the IVT based method yielded higher amounts of amplified RNA and more reproducible results.

A study from Saghizadeh (Saghizadeh *et al.*, 2003) showed that the relative abundance of RNA species in amplified samples is maintained after amplification. This conservation is of course an absolute prerequisite for subsequent expression profiling experiments.

As higher quantities (starting from 5 to 50 μg depending on protocols and arrays for replication of measurements) are required to obtain meaningful and reproducible signals, an amplification protocol was established and used in this work based on *in-vitro* transcription (Van Gelder *et al.*, 1990). If using IVT it has to be considered that due to the applied method, only the sequences close to the poly-A tail of the mRNA are amplified. This leads to smaller fragment sizes of aRNA compared to mRNA. However this effect does not have significant influence on the outcome of microarray experiments, especially when the probes on the array are carefully selected and only one round of amplification is required as in this study. When working with microdissected samples where a second round of amplification is needed due to the low amounts of starting mRNA an accumulation of shortened products after two amplification rounds might occur. If the shortened sequences lack their complementary part which is used on the array the transcription profile is altered. To conclude IVT is not free of problems as it amplifies templates that are rare, GC rich or show strong secondary structures inefficiently. While it seems to increase the number of genes that can be detected in small samples, it has also been reported to decrease the number of differentially expressed transcripts found by 20 to 30 % (Xiang *et al.*, 2003). This is however an acceptable price for being able to study more homogeneous and informative samples from primary cell cultures.

In this study one round of amplification based on IVT yielded 100 μg of aRNA from 1 μg of total RNA. Under the assumption that about 1 to 5 % of total RNA is poly A RNA, this corresponds to a 2,000 to 10,000-fold amplification. During the amplification biotin labeled nucleotides were incorporated into the aRNA. The incorporation of biotinylated adenosine and cytosine did not affect amplification in concern of product yields.

From the results obtained in this work it can be concluded that the techniques used for RNA purification are well understood, optimized and validated. The different kits for RNA extraction compared in this work, yielded all high amounts of not degraded RNA.

As quality and quantity was more or less the same, the choice of the applied method is based on price and convenience. As none of the extraction methods is able to create intact RNA out of degraded one it is important to determine the quality of RNA in advance of amplifying and/or labeling. Decidedly the best choice is using the Agilent Bioanalyzer 2100 to assess quality of RNA samples. Also success of amplification can be determined by using this capillary gel electrophoresis based assay. To study expression profiles from small template amounts, template or signal amplification is essential. Template amplification has been used in this work and also in many other laboratories. Among the many methods used to amplify RNA, IVT by phage RNA polymerase seems to be the most robust but also time-consuming and expensive.

4.1.3 Comparison of Expression Data from Different Microarray Platforms

Comparison of different microarray platforms, especially when used to elucidate similar questions is of considerable interest. The question is whether gene expression data obtained from different platforms is comparable and will lead to similar conclusions. Several studies especially between 2000 and 2006 have addressed this question. However there is still no clear answer how identical results are and have to be. This conclusion can be found in the published data as well as a results obtained in this study. The opinions and published results are as heterogeneous as the expression data itself. In 2003 three microarray platforms were compared with a more or less poor result (Tan *et al.*, 2003). Absolute and relative gene expression measurements on arrays from Affymetrix, Agilent and Amersham were compared and gave moderate correlation. Affymetrix and Amersham based on short oligonucleotides and biotin labeling correlated with a factor of 0.52. This result was in the same range as the comparison with cDNA based chips from Agilent. The correlation between Agilent and Affymetrix was 0.53 and with Amersham 0.59. In the same year Barczak found moderate correlation as well, comparing Affymetrix chips and spotted 70mer oligonucleotide arrays (Barczak *et al.*, 2003). Signal intensities showed correlations between 0.56 and 0.60. However a strong correlation was observed for relative expression ratios (0.8). One year later a comparison between Affymetrix GeneChip and Amersham CodeLink arrays could be remarkably improved by removing genes within platform noise. The

correlation was first 0.62 and after omitting genes in the noise it was 0.79. In the same year a comparison between Affymetrix and cDNA arrays was published (Mah *et al.*, 2004). The absolute expression levels obtained with the fluorescence based short oligonucleotide arrays from Affymetrix and the radioactive labeled cDNA arrays showed a poor correlation. Correlation of gene expression values determined with Affymetrix short oligonucleotide arrays, Agilent cDNA and custom made cDNA arrays was good in the work from Jarvinen and his co-workers (Jarvinen *et al.*, 2004). The correlation between Affymetrix and the cDNA based arrays was 0.8 in this study. This was in contrast to the study of Tan *et al.*, 2003. In May 2005 several new studies were published in Nature Methods dealing with the comparison of different microarray platforms. Larkin *et al.*, 2005 compared gene expression between two microarray platforms: the short oligonucleotide Affymetrix Mouse Genome 430 2.0 GeneChip and a spotted cDNA array using a mouse model of angiotensin II-induced hypertension. They found that the experimental treatment had a greater impact on the measured expression than the used platform. This observation was true for more than 90 % of the analysed genes and was validated using qRT-PCR. In another study (Irizarry *et al.*, 2005) a multi laboratory comparison of Affymetrix GeneChips, spotted cDNA and oligo nucleotide arrays was performed. Irizarry and co-workers demonstrated in this study that it was more important where the array experiment was done than on which platform. The results were more affected by the laboratory than by the different applied methods. Bammler *et al.*, 2005 investigated sources of error and data variability between laboratories and across microarray platforms. RNA expression data were generated in seven laboratories, which compared two standard RNA samples using 12 microarray platforms. Reproducibility for most platforms within any laboratory was typically good, but reproducibility between platforms and across laboratories was generally poor. Reproducibility between laboratories increased markedly when standardized protocols were implemented for RNA labeling, hybridization, microarray processing, data acquisition and data normalization. As a conclusion it was found that microarray results can be comparable across multiple laboratories, especially when a common platform and set of procedures are used (Bammler *et al.*, 2005). In 2006 the MicroArray Quality Control (MAQC) project was published in Nature Biotechnology dealing with inter- and intraplatform reproducibility of gene expression measurements.

This extensive study comprising 1,327 microarrays addressed questions about the reliability of microarray technology, as previous studies with dissimilar or contradictory results raised concerns about reliability and reproducibility. The main study compared seven different microarray platforms and included 60 hybridizations per platform. The study furthermore captured intrasite, intersite and interplatform differences. In the MAQC study Affymetrix, Agilent and Illumina displayed a high correlation value of 0.9 with TaqMan assays based on a comparison of approximately 500 genes. The gene list of differentially expressed genes showed about 89 % overlap between different laboratories using the same platform. Microarray results were generally repeatable within a test site and a high correlation between different platforms was observed. However it should be noted that the good correlating results from this study were obtained by comparing two sample types with the largest difference among the four sample types used in the MAQC project. In practical applications, the expected differences between sample types (induced versus control) are usually much smaller compared to those in this study. Therefore the high comparability of microarray data shown in this MAQC study does not reflect the achieved consistency of data obtained in studies dealing with real samples. This difference was also found by comparing the published comparison with the consistency of data in this study presented here. However the overlap of differentially expressed genes determined with different platforms was still good.

The reproducibility of microarray experiments is relying on three major factors: biological variability (based on different samples) as well as technical variance (cultivation, sample preparation, labeling and hybridization) and as an additional source of variation post-laboratory steps of scanning and quantification (e.g. strong background signals, dust, different scanners/settings, old lasers) as well as different methods of data analysis and interpretation (Churchill, 2002).

The biological variation can be influenced by genetic- or environmental factors. Both but especially the last one are very different in the human population compared to experiments with mice or microorganisms. Living habitats like smoking, alcohol consumption, nutrition, degree of fitness, and many others have strong impact on the gene expression level and therefore on the comparability of different biological replicates. The previous history and indication of the patients as well as the treatment of

the disease have strong influence on gene expression. All the mentioned challenges can be depleted when working with microorganisms or animal experiments e.g. with mice. However the obtained results from an experiment with an animal model can not be assigned to the situation in humans. To obtain data reflecting the human situation experiments with human cells are inevitable.

Even if the genetic and environmental background is similar, like in experiments with animals or microorganisms, the differences based on technical variation and measuring errors remain. The technical variation is based on differences in handling during sample generation and processing. This kind of variation can be determined by analysing the same biological samples several times.

To determine the effect of treatment in different biological samples the statistical test has to be based on the biological variance. If variations in treatment groups are the main interest the test has to be based on technical variation. Further more it is important to replicate biological samples in order to draw conclusions that are valid beyond the scope of the particular samples investigated. In this study six biological replicates were analysed. Samples from three different donors were analysed with full genome arrays from Affymetrix and the differential expression was determined for additional three patients with sub-genome arrays.

The intraplatform data repeatability and reproducibility determined in this study was good. The correlation between similar hybridizations on Affymetrix GeneChips was 0.93. A comparable correlation for the Affymetrix platform had been reported previously by Baugh *et al.*, 2001. In this study the correlation between four hybridizations was 0.99. The Eppendorf platform provided also highly reproducible measurements (0.8) of gene expression profiles.

The interplatform comparability from the work presented here showed similar results to previous publications. Expression values generated on different platforms can not be compared directly with each other, as unique pre-processing methods and probe sequences result in variable signals for the same target. To overcome this problem relative expression of samples between different platforms shall give similar results. The log ratio of gene expression between two sample types (Rifampicin and DMSO) was compared. Furthermore the obtained gene lists showing the differentially expressed genes for each platform were investigated. It has to taken into consideration that each

microarray platform has its own background correction and dynamic range of signal detection, which can lead to over- or underestimation of log ratios or fold changes. Nevertheless a correlation of 0.6 was obtained between data from Affymetrix, Agilent and Eppendorf. This observation was in agreement with previously published studies.

4.1.4 Network Reconstruction Based on Microarray- and Literature Data

Both literature research and microarray based approaches share the common goal of identifying the hidden networks of biological components. Both the experimental data and the literature knowledge were both iterative integrated in the work presented here. This approach of biological network modeling seems to be an effective way to recover transcriptional regulatory networks from static microarray data (Le Phillip *et al.*, 2004). Incorporating prior knowledge into the learning scheme greatly reduces the data required. The network reconstruction in the work presented here focused on finding interactions and connections of differentially expressed genes co-cited in literature (literature mining and microarray analysing (LMMA-) approach). This approach will unravel so far in this context undiscovered but known (published) interactions. In addition the reverse engineering techniques, used at the Institute of Biochemical Engineering to find new regulatory interactions and interaction partners from the microarray data sets, will reveal completely new biological knowledge. Based on the snapshots of the transcriptome, determined with microarrays determining the relative expression levels of genes at a genome wide level, it is possible by applying algorithmic approaches to recover regulatory networks (de Jong, 2002; van Someren *et al.*, 2002). The main focus of the reconstruction was put on a small regulatory sub-network of CYP3A4 induction.

4.2 Physiological Considerations

The main part of the physiological analysis was the regulation of CYP3A4 by PXR and other regulatory proteins after stimulation with Rifampicin. The metabolism of Rifampicin by deacetylation and hydrolysis is shown in Figure 118. It was however

more important that Rifampicin is a good inducer of several genes involved in xenobiotic metabolism which was a prerequisite for studying regulatory principles based on perturbation of liver cells.

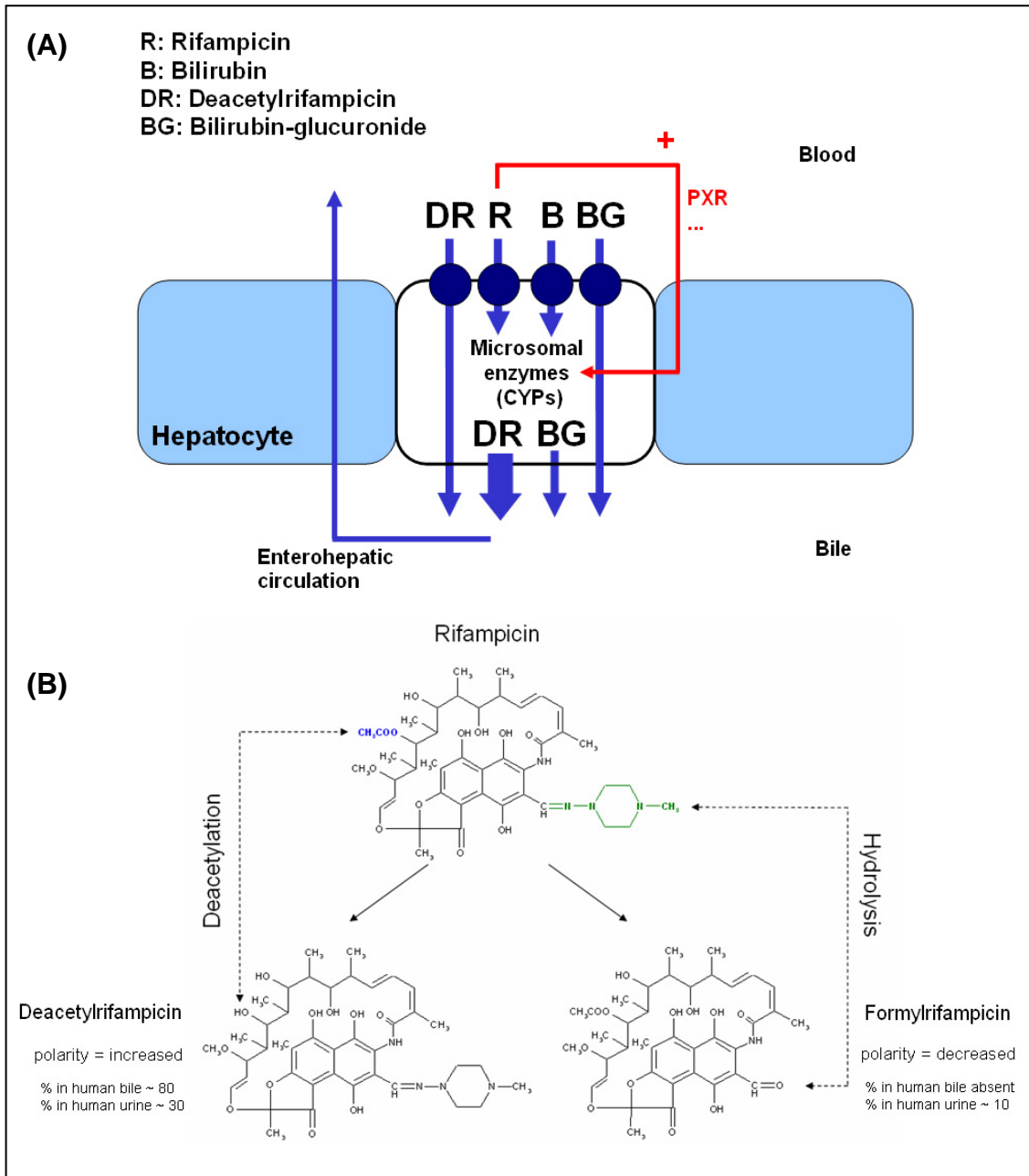


Figure 118: Hepatic metabolism of Rifampicin (R), Bilirubin (B), Deacetylriofampicin (DR), and Bilirubin-glucuronide (BG). (A) Eighty per cent of oral Rifampicin enters the enterohepatic circulation. Some Rifampicin passes through the hepatocyte unchanged, but most is deacetylated before excretion into bile. Some of the Deacetylriofampicin is further glucuronidated. Graph based on Johnson and Pavord, 1996. (B) Main metabolic derivatives of Rifampicin in man, polarity and percentage of recovery in bile and urine based on Acocella, 1978.

During investigation of the regulatory network, which is still the most important part of this work, it became clear that the regulation of genes involved in xenobiotic metabolism can not be investigated without taking interaction with other circuits into account. An overview of genes affected by the perturbation and the involved metabolic routes and regulatory circuits is shown in (Figure 119 and Figure 120). In the following part it is shown that e.g. the expression of aminolevulinic synthase relies on the same regulatory mechanisms as CYP3A4. Furthermore the influence of the Rifampicin induction on the expression of genes involved in bile acid-, glucose- and fat metabolism as well as cell adhesion and oxidative stress is briefly analysed.

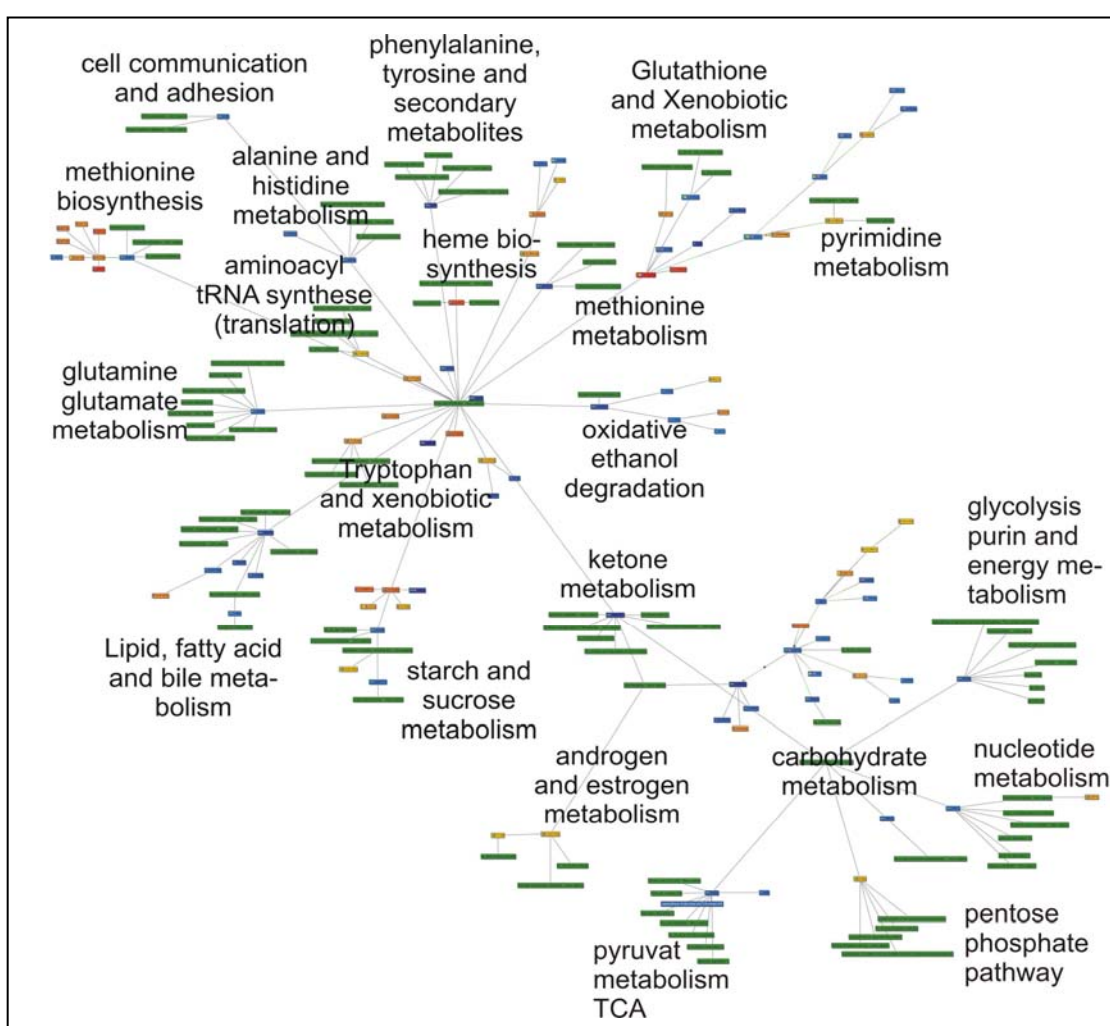


Figure 119: Metabolic pathways containing genes found to be differentially expressed after stimulating human hepatocyte cells with Rifampicin. Next to the xenobiotic metabolism a lot of other metabolic functions are perturbed by induction with this drug. The green bars stand for a metabolic function affected while the small squares represent genes which were found differentially expressed. Blue colour shows down-regulation while red stands for up-regulation.

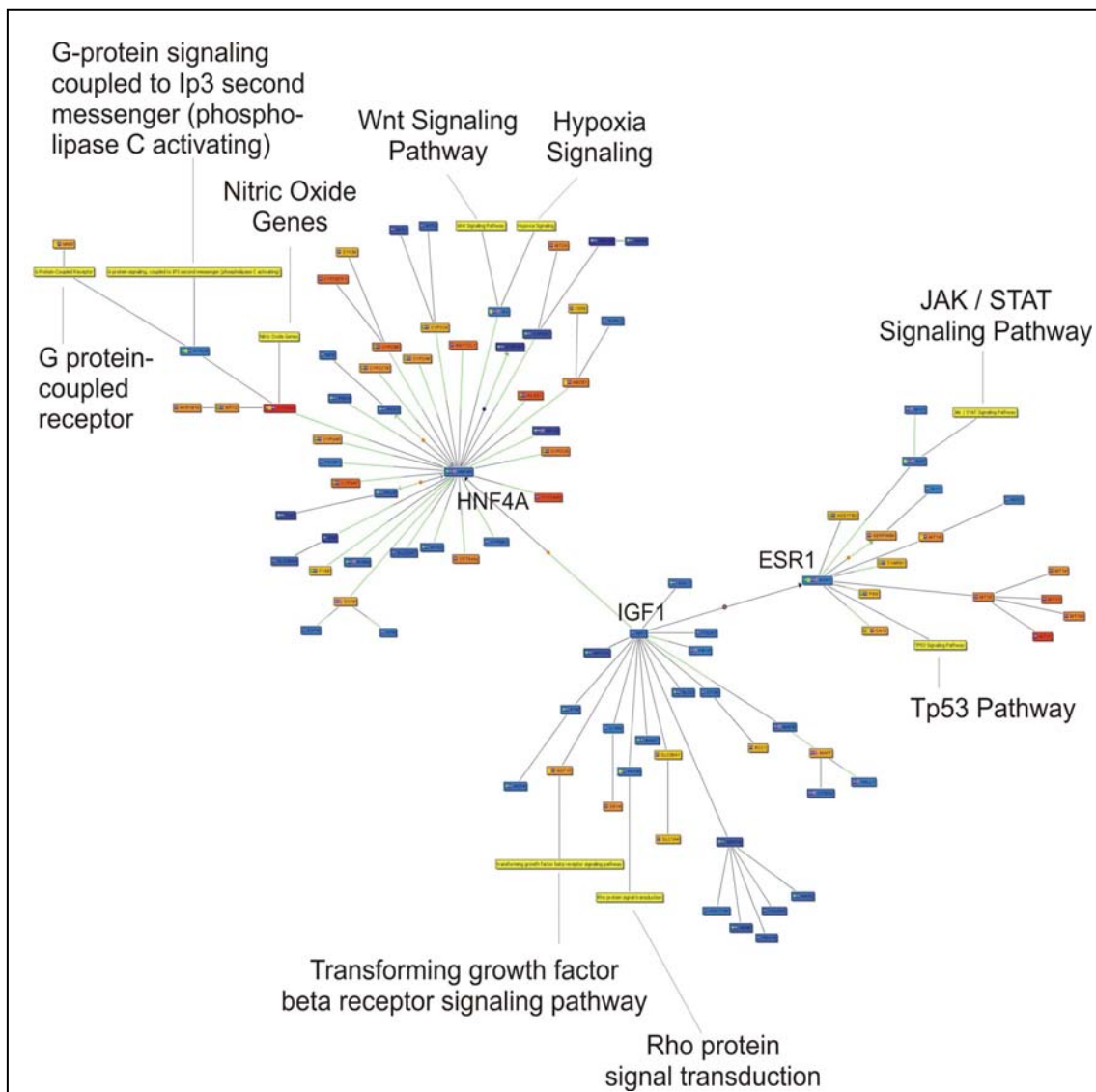


Figure 120: Regulatory pathways containing genes found to be differentially expressed after stimulating human hepatocyte cell cultures with Rifampicin. The centres of regulation HNF4A and IGF1 were found to be down-regulated. Further affected were stimulatory and inhibitory G proteins which allow the cells to accurately regulate their metabolism. The G protein, or heterotrimeric guanine nucleotide-binding regulatory protein, acts as a versatile messenger for transducing signals for a variety of hormones that cause similar cellular responses such as the activation of cAMP production. The Wnt signaling pathway describes a complex network of proteins most well known for their roles in embryogenesis and cancer, but also involved in normal physiological processes in adults (Lie *et al.*, 2005). Hypoxia and oxidative stress, often tumour related, may trigger hypoxia signaling. Also correlated with tumours is the TP53 (tumor suppressor 53) pathway. The JAK-STAT signaling pathway takes part in the regulation of cellular responses to cytokines and growth factors. Transforming growth factor beta (TGF β) receptor signaling controls many cellular functions like proliferation and differentiation. Rho GTPases have been shown to regulate many aspects of intracellular actin dynamics and further have been described as ‘molecular switches’ playing a role in cell proliferation, apoptosis, cell division, gene expression and multiple other common cellular functions (Boureux *et al.*, 2007; Bustelo *et al.*, 2007). The regulation of xenobiotic metabolism in general seems to be more based on posttranslational modification (phosphorylation, etc.) and binding events (like TFs to the promoters of genes) instead of transcriptional change.

4.2.1 Control of Cytochrome P450 Enzyme Family (CYPs) and 5-Aminolevulinic Synthase (ALAS1) Gene Transcription by Nuclear Receptors

The nuclear receptor superfamily is a group of transcription factors which regulates genes involved in a variety of physiological, developmental, and metabolic processes. Members of this superfamily play a role in regulation of detoxification, reproductive systems by steroid hormones, control of development and regulation of bile acid and cholesterol biosynthesis (Figure 121).

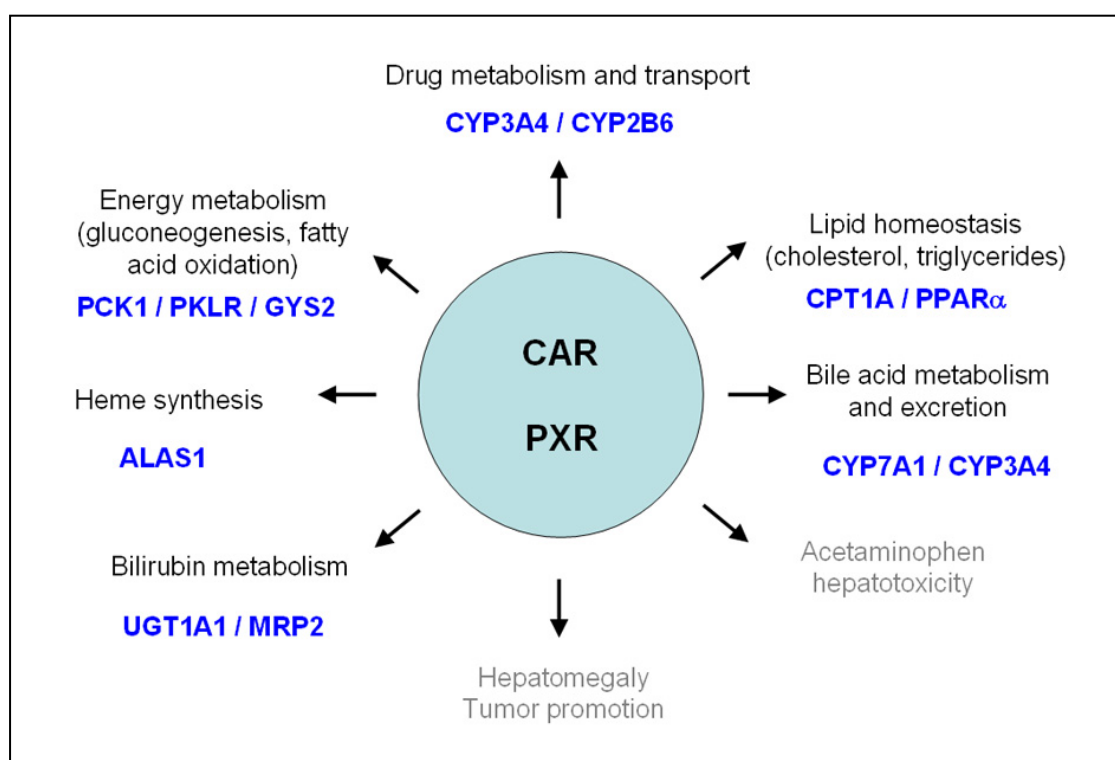


Figure 121: Versatile physiological roles of CAR and PXR. Investigated biological functions regulated by CAR and PXR. Shown in blue are genes found to be differentially expressed in this study.

In addition to their normal function, loss of control of nuclear receptor signalling pathways contribute to the development of endocrine-related diseases, such as breast cancer, prostate cancer, ovarian cancer, diabetes and obesity (<http://www.nursa.org>).

The formation of xenobiotic metabolizing cytochrome P450 enzymes (CYP) is based on the interaction of nuclear receptors and co-activators. These DNA-binding proteins transform their activation by ligands like lipophilic drugs, solvents, alcohol or steroids into an activation of gene transcription. The CAR/PXR mediated activation is enhanced

and modulated by several co-activators which are parts of other genomic circuits. The expression of target genes of the CAR/PXR regulation is increased e.g. by the binding of glucocorticoid receptor (GR) activated by a steroid hormone. This could be e.g. cortisol generated in fasting. Not only the genetic background and the induction with Rifampicin itself have influence on the expression. Factors like nutrition, stress, infections or other might have strong effects on the expression of genes, as they trigger genetic circuits interacting with the detoxification network.

In response to demands for cytochrome P450 (CYP) species activation of the *cyp* genes takes place concertedly with the gene for the enzyme delta-aminolevulinic acid synthase 1 (ALAS1) as shown in this study. Intensified metabolism through activation of heme oxygenases will decrease the amount of available heme and therefore increase the demand for heme synthesis. A synchronously expression of *cyp* genes and *alas1* assures the supply of needed heme for CYP formation. ALAS1 is the transcriptional regulated rate-limiting enzyme of the heme biosynthesis (Figure 122).

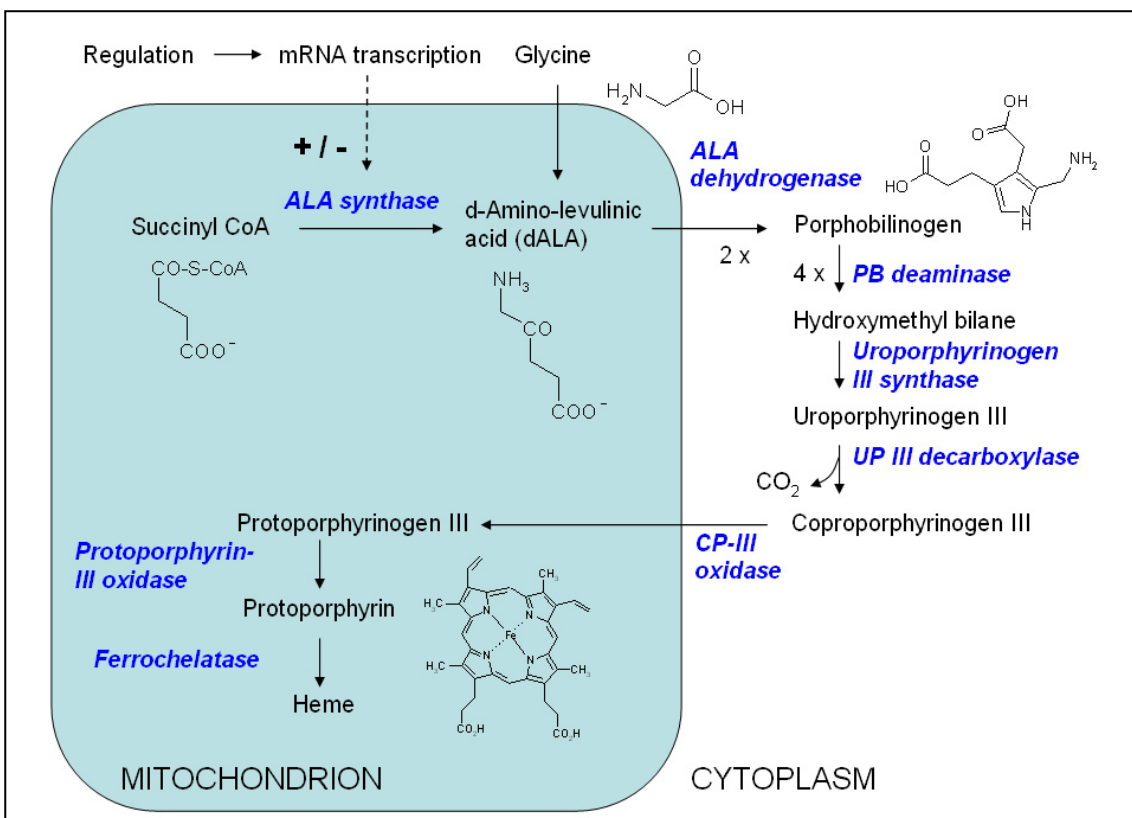


Figure 122: Heme biosynthesis pathway. The gate-keeping inducible enzyme 5-amino-levulinic acid synthase (ALAS) initiates biosynthesis of heme by condensing succinyl CoA and glycine followed by seven sequentially catalyzed reactions. The enzymes are given in blue. ALAS is the only enzyme of all eight enzymes in this pathway to be transcriptionally regulated.

Heme is an essential component of numerous heme-proteins with functions including oxygen transport, energy metabolism, and drug biotransformation. Upon exposure to drugs that induce cytochrome P450s and other drug-metabolizing enzymes, aminolevulinic acid synthase (ALAS1) was found to be transcriptionally up-regulated in this study, increasing the rate of heme biosynthesis to provide heme for cytochrome P450 heme-proteins (Podvinec *et al.*, 2004). The first step of the heme biosynthetic pathway is catalyzed by this aminolevulinic acid synthase (ALAS). The enzyme requires glycine and succinyl coenzyme A (CoA) as substrates, and pyridoxal 5'-phosphate as a cofactor. The synthesis of heme is regulated at the initial step on the level of transcription of *alas* gene.

Drugs effecting CYP-induction will consequently activate transcription of ALAS1. To achieve co-regulation, the activation of transcription of *cyp* genes and *alas* follow the same mechanisms based on association of transcription factors (Figure 123).

Aminolevulinic acid synthase (ALAS1) and members of the Cytochrome P450 enzyme family (CYPs) are regulated by similar mechanisms based on CAR, PXR, and RXR. As promoter induction for *cyp* genes and *alas* share similar binding sites for at least three co-activators, enhancing nuclear receptor mediated transduction, it is supposed that an equivalent formation of ALAS1 and CYP mRNA is desired. Nuclear factors known to enhance the CAR/PXR transduction are the ligand independent growth hormone controlled hepatocyte nuclear factor 4 (HNF4), the insulin responsive forkhead box class O- (FoxO) protein and the proliferator activated receptor gamma co- activator 1 alpha (PGC-1-alpha). FoxO is activated as a response to different kind of stress and PGC-1-a is responding to glucagon liberation in fasting. These many interactions and cross-talks between different circuits might be one reason for the large inter- and intra-individual variability in response to a specific induction, e.g. with Rifampicin.

The ALAS1 induction based on a Rifampicin stimulus observed in this study after 24 h of induction is generated in response to demands for heme for the supply of catalytic cytochrome P450s. These CYPs are the main player of detoxification in phase I metabolism of xenobiotics. Most important are the members CYP2 and 3A4 which are the most dominant human hepatic heme proteins.

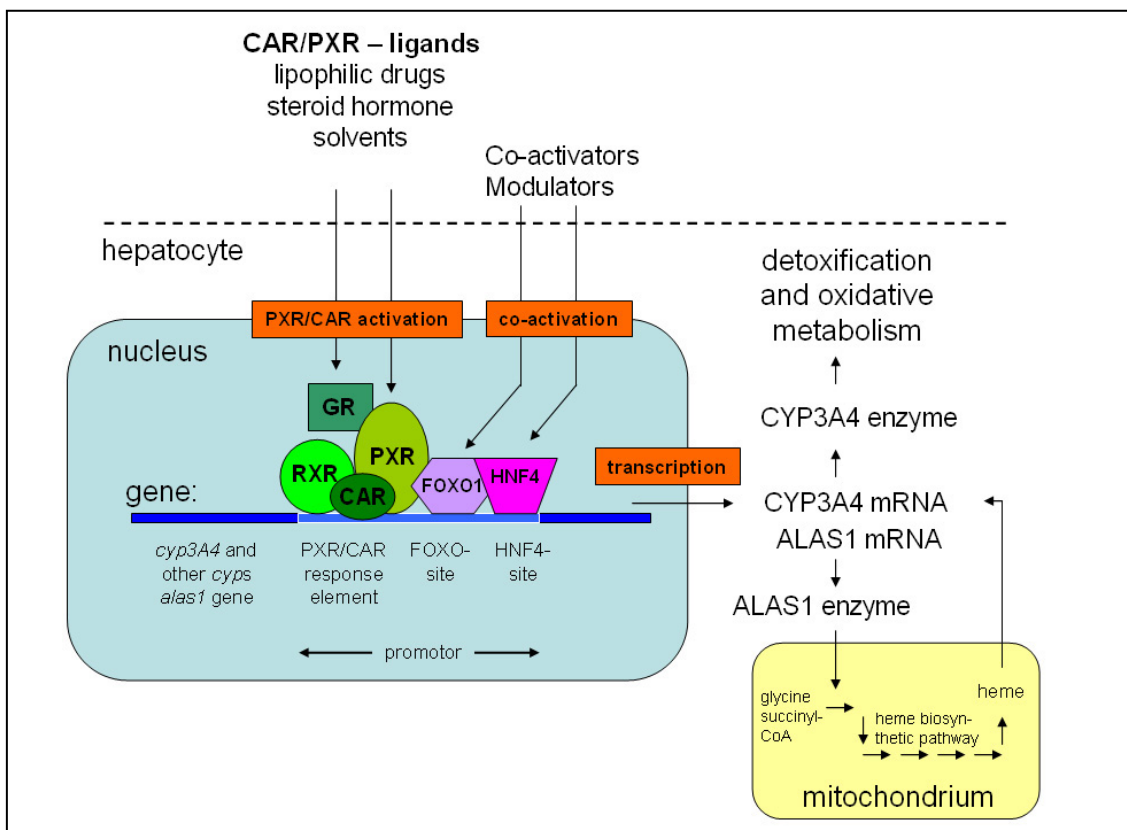


Figure 123: Regulation of Cytochrome P450 enzymes (CYPs) and aminolevulinic acid synthase (ALAS1) expression based on binding events of nuclear receptors and co-activators. Transcription of CYP and ALAS1 genes takes place synchronously as ALAS1 controls production of heme needed for formation of the active centre of CYPs. The nuclear receptors regulating this transcription are activated by exogenous and endogenous ligands like xenobiotics or steroid hormones. Co-activation of the gene can occur by binding of activated transcription factors FoxO1 and HNF4 to specific sites of the promoter region. The activation and binding events lead to equivalent formation of ALAS1- and CYP-mRNAs.

There are several mechanisms of induction of xenobiotic metabolising CYPs suggested. While most of these induction mechanisms are based on increased rates of transcription, regulated by interaction of receptor proteins, another mechanism also focuses on an increased stability of the formed enzyme. Such a post-transcriptional stabilisation is suggested for CYP2E1 (Koop *et al.*, 1985; Song *et al.*, 1989; Moncion *et al.*, 2002). However, the most important way of increasing the amount of active CYP enzyme in a cell is by up-regulated transcription, based on receptor interaction in the hepatocyte nuclear compartment. As the physiological ligands were unknown, these receptors were called orphan nuclear receptors (NRs). By binding their receptor specific ligands, the

NRs become activated. These ligands can be xenobiotics, like drugs, alcohol, solvents or endobiotics like hormonal steroids. The binding of 9-cis retinoic acid xenobiotic receptor (RXR) to the heterodimerization complex initiates the attachment to recognition sites within enhancer sequences of their target genes. This binding starts the concerted transcription of CYP and ALAS1. The modulators of transcription of *cyp* genes belong to the growth hormone (GH) network of liver transcription factors. This growth hormone based activation might be one reason for sex dependent CYP regulation. Differences between males and females can be explained by sex dependent control of growth hormone activity and glucocorticoid receptor mediated interaction of CAR/PXR activation.

In humans the constitutively active receptor (CAR) and the pregnane xenobiotic receptor (PXR) regulate most of the xenobiotic induced CYP transcriptional responses. The majority of all presently prescribed drugs (> 50 %) are metabolized by target genes activated via CAR/PXR based mechanisms. Few prescription drugs are ligands of the aromatic hydrocarbon receptor (Ah-receptor), which is said to be mainly activated by polycyclic hydrocarbons. Another way of induction is via the peroxisome proliferator activated family of receptors (PPARs), which play fundamental physiological roles in the regulation of energy balance as shown in this work by the Rifampicin induced change in expression of PPARs interfering with the fatty acid metabolism.

The network and interactions between nuclear receptors like CAR and PXR and their co-activators are very complex. Information about these interactions is crucial for a successful modelling of the detoxification process. Not only the identity and number of interaction partners are of interest but also the effects of endogenous and exogenous factors able to modulate activation of gene transcription. The mechanisms involved have been studied over the last few years and a good overview is given in the following papers (Waxman, 1999; Dussault and Forman, 2002; Wang and LeCluyse, 2003; Honkakoski *et al.*, 2003; Tompkins and Wallace, 2007). Some important aspects on the action of nuclear receptors will be given in detail in the following section.

At the beginning of the regulatory circuit leading to a transcriptional response to the exposure stands the binding of a ligand (steroids, drugs, xenobiotics like Rifampicin) to the ligand binding domain of the nuclear receptor. To allow binding of lipid soluble ligands in different orientations and different sizes the ligand binding domain is

hydrophobic and highly flexible. The large, smooth shape of the binding pockets of CAR and PXR allow a large number of substances to induce CYPs. The only common property is their lipophilic nature. Based on the degree of fit to the pocket, ligands have different transactivation potencies. Each inducer has its own CYP-inducing ability and inducing power. The amount of available nuclear receptors is regulated by ligand binding based glucocorticoid receptor activity. In response to binding of the ligand GR increases the expression of nuclear receptors. Many different proteins further modulate the expression of target genes as co-activator or co-repressor by binding to the promotor regions of these genes.

The second step of the induction process is binding of the nuclear receptor ligand complex to the target genes. Several different response elements were recognized in the *cyp* genes. Each response element has a different nuclear receptor specificity and response activity. Therefore the transcriptional response is quite different between inducing ligands.

The most abundant inducible hepatic CYPs belong to the CYP2 and 3A families which are activated by CAR/PXR. The same regulatory mechanism lead to transcription of ALAS1 enzyme needed for biosynthesis of the catalytic heme component of the CYPs. The CYP2 genes contain a phenobarbital response element (PBREM) with two NR binding sites based on a direct repeat (DR) 4 motif. At these binding sites the heterodimerization complex of drug activated CAR and RXR can bind. Under not induced conditions CAR is found in the cytosol. The strength of CYP-inducing abilities depends on the degree of translocation into the nucleus. The mechanism of the translocation is not known so far. However proteins preventing the translocation of CAR into the nucleus like the cytoplasmic CAR retention protein (CCRP, Dnajc7, Hsp40) have been identified in rats. Peroxisome proliferators like nuclear receptor peroxisome proliferator-activated receptor α (PPAR α) activate and enhance the nuclear translocation of CAR in mouse liver cells in vivo (Guo *et al.*, 2006; Sarkar *et al.*, 2007). Furthermore phosphorylation and dephosphorylation seems to be quite important in preparing CAR for translocation into the nucleus. This drug activated phosphatases are another important factor for CAR mediated CYP induction.

CYP3A4 shown to be strongly up-regulated in this study in Rifampicin induced cells compared to DMSO control cells, has two DR-3 and ER-6 motives in the promotor

region. Ligand activated RXR-PXR as well as CAR-RXR dimers will bind to these sites increasing transcription. While CAR has to be translocated into the nucleus PXR seems to reside in the nucleus. Although this question is not finally answered PXR activation seems not to require any translocation to access target genes. The rate limiting step of activation in this case would be the binding of ligand instead of translocation. The variable induction is due to different affinities between ligand and receptor and intranuclear ligand concentrations.

For the co-regulation of ALAS1 with CYP3A4 the *alas1* gene contains also a DR4 recognition site for PXR and CAR (Podvinec *et al.*, 2004). The expression of ALAS1 is induced by a large number of drugs. As this induction relies on the action of the same nuclear receptors inducing CYP transcription, a coordinated response to xenobiotic stress and endogenous metabolic tasks is guaranteed.

Fine-tuning of induction by PXR and CAR is achieved using several co-activators and co-repressors binding to the activation function domains (AF). Sites for nuclear-receptor linked transcriptional modulation by e.g. AF-, COUP- and the CCAAT/enhancer binding protein alpha (C/EBPalpha)-binding sites can be found in CYP promoters as well as in *alas1* gene.

The transcription factor C/EBPalpha regulates a number of liver stress-response genes and participates in the regulation of CYP3A genes by glucocorticoids and inflammation (IL-6). Single nucleotide polymorphisms (SNP) in the C/EBPalpha binding sites of CYP3A promoters abolishes the activation via this route (Podvinec *et al.*, 2004; Rodrigues *et al.*, 2003).

Modulation of PXR/CAR transactivation occurs via several routes shown in yellow in Figure 124. The regulation machinery involved in the induction of CYPs and ALAS1 is altered by a growth hormone pulse via hepatocyte nuclear factor 4 (HNF4), via the glucocorticoid receptor (GR) activated by the so called hypothalamic-pituitary-adrenal (HPA) signalling pathway starting from the central nervous system (hypothalamus) over several hormones (CRH, ACTH) to the excretion of corticosteroid hormones. Two other ways are via insulin responsive Forkhead box class O (FoxO) protein activity and via the non-insulin dependent PGC-1 alpha pathway activated by glucagon.

others the most important one interacting with HNF4 are steroid receptor co-activator 1 (SCR1) and p300 / CCAAT box/enhancer binding protein (p300 / c/EBP) (Wang *et al.*, 1998; Wang *et al.*, 2006) as well as peroxisome proliferator-activated receptor gamma co-activator 1alpha (PGC-1alpha) and forkhead box O (FoxO) which is activated in fasting (Handschin *et al.*, 2005). In calorie restriction or starvation, FoxOs are transcriptionally active in the nucleus and decrease insulin secretion, increase hepatic glucose production and food intake and cause degradation of skeletal muscle for supplying substrates for glucose production (Nakae *et al.*, 2007). A complex between HNF4 and PGC-1alpha controls gluconeogenesis via transcription of the genes for phosphoenolpyruvate carboxykinase (PEPCK) and glucose-6-phosphatase (Glc6P). Ligand activated PXR separates HNF4 and PGC-1alpha from each other resulting in a down regulation of hepatic glucose production even under fasting.

By co-activating PXR, HNF4 controls bile acid biosynthesis via transcription of the rate limiting enzymes in the production of bile acids from cholesterol: cholesterol 7alpha hydroxylase (CYP7A1) and sterol 12alpha hydroxylase (CYP8B1) (Bhalla *et al.*, 2004). Through feed-back control, bile acid precursors down-regulate PXR via activation of the farnesoid xenobiotic receptor (FXR), a nuclear receptor that up-regulates the PXR transactivation repressor small heterodimer partner (SHP) (Chen *et al.*, 2001; Eloranta and Kullak-Ublick, 2005). These two effects on bile acid metabolism and glucose homeostasis will be discussed in separate chapters (4.2.2; 4.2.3).

Under different forms of stress induction of CYPs and ALAS1 is achieved via activation of transcription factors belonging to the FoxO family of proteins. Environmental and metabolic stimuli are translated into changes in gene expression by these transcription factors. Regulatory functions of FoxO proteins are induced by post-translational modifications like phosphorylation or acetylation. Based on these modifications the proteins are directed to their targets, change binding affinities and interaction partners to fulfil their divergent tasks. Targets for FoxOs are genes controlling hepatic generation of glucose as well as production of CYPs and ALAS1 (Barthel *et al.*, 2005). The localisation and activation of these enzymes is regulated by glucagon and insulin the key hormones of glucose homeostasis.

Another important co-activator of several nuclear receptors is PGC1alpha. It is induced in the liver by glucagon during fasting and extensive physical practice. Glucagon is

together with glucocorticoids a signal of reduced plasma glucose levels. PGC1alpha modulates the co-expression of CYP, ALAS1, PEPCK and GLC-6-P. Next to the functional role in adjusting the expression of genes involved in xenobiotic metabolism this co-activator is a second way, next to the FoxO route, of hepatic glucose liberation and gluconeogenesis.

Glucocorticoids as part of the hormone mediated regulation play an important part in the cellular response to metabolic stress. The name glucocorticoid is derived from observations that these hormones are involved in glucose metabolism. Cortisol stimulates several processes maintaining normal concentrations of glucose in blood. In the liver these effects include regulation of gluconeogenesis resulting in the synthesis of glucose from non-hexose substrates like amino acids and lipids. In this study a down-regulation of genes involved in gluconeogenesis was observed. By activation of the glucocorticoid receptor (GR) the glucocorticoids constitute to the whole regulation through effecting transcription of many genes (Pascussi *et al.*, 2001; Falkner *et al.*, 2007). Activation of CYP2 und CYP3A4 by ligand activated CAR/PXR is enhanced by GR. In addition the liver transcription factor C/EBPalpha, known to regulate a number of other stress response genes, takes part in the transcriptional control of CYPs via GR (Rodrigues *et al.*, 2003). The regulation of CYPs via the central nervous system is of high importance especially when considering pharmacological drug administration stimulating or inhibiting the brain dopaminergic system (e.g. by dopamine receptor-blocking neuroleptics). Interference with CYP expression might cause pharmacokinetic changes leading to a reduced or enhanced efficacy of drugs due to altered metabolism and detoxification (Wojcikowski *et al.*, 2007). Glucocorticoids in combination with HNF4alpha are further able to synchronize the liver circadian transcriptome leading to a coordinate transcription of tissue specific metabolic pathways (Reddy *et al.*, 2007).

Glucocorticoids have two different roles in CYP3A4 induction. Under normal physiological conditions at low micromolar range a xenobiotic independent CYP3A4 induction is given based on the interaction of PXR and CAR initiated by GR which require c/EBPalpha. During experimental or pharmacological stress, at high micromolar concentrations, glucocorticoids lead to a strong induction of CYP3A4 by direct ligand activation of PXR (Pascussi *et al.*, 2001; Wang *et al.*, 2003; Wang and LeCluyse, 2003). Further important effects on the HPA axis are e.g. the remarkable infection and

inflammation based down-regulation of hepatic drug metabolism during inflammation which has been known for several years. The mechanisms behind this effect have not been fully elucidated. The down-regulation of CYPs might be based on pro-inflammatory cytokines such as IL-6 (also observed in this study; Teng and Piquette-Miller, 2005). The inflammation linked regulation is discussed separately in 4.2.6.

There is a complex interplay among the different regulatory circuits controlling the transcription of CYPs and ALAS1. The key for the activity of nuclear receptor based regulation of gene transcription, which was investigated using DNA microarrays in the work presented here, seems to be the interaction and binding of nuclear receptors and co-activators or co-repressors. These crucial interactions between ligand and receptor as well as between receptors and modulating molecules are further elucidated in the second phase of the Hepatosys project at the Institute of Technical Biochemistry using Biacore measurements.

The available molecular data on functional interactions between nuclear receptors and co-regulators has exciting implications for the development of novel pharmaceutical therapies for a wide range of diseases. As an example clinical strategies addressing the role of co-activators and co-repressors as toolbox for initiating stimuli for cell proliferation regulated by nuclear receptors will succeed in treatment of breast cancer and other reproductive cancers. Also the involvement of co-regulators and nuclear receptors in energy metabolism suggest better treatments of nutritional disorders and a rewarding way of research.

4.2.2 Regulation of Bile Acid Metabolism by Nuclear Receptors

Bile secretion plays a major important role in liver physiology because it serves as the excretory route for many endo- and xenobiotics like bile acids, bilirubin, cholesterol, phospholipids and drugs (Zollner and Trauner, 2006). Bile acids are not only indispensable for elimination of excess endogenous substances like cholesterol, but play also an important role in the absorption of dietary lipids and fat soluble vitamins. However, bile acids are detergents that can become extremely toxic if their levels become abnormally high (Kliwer and Willson, 2002) making a sophisticated regulation of bile acid levels essential (Figure 125).

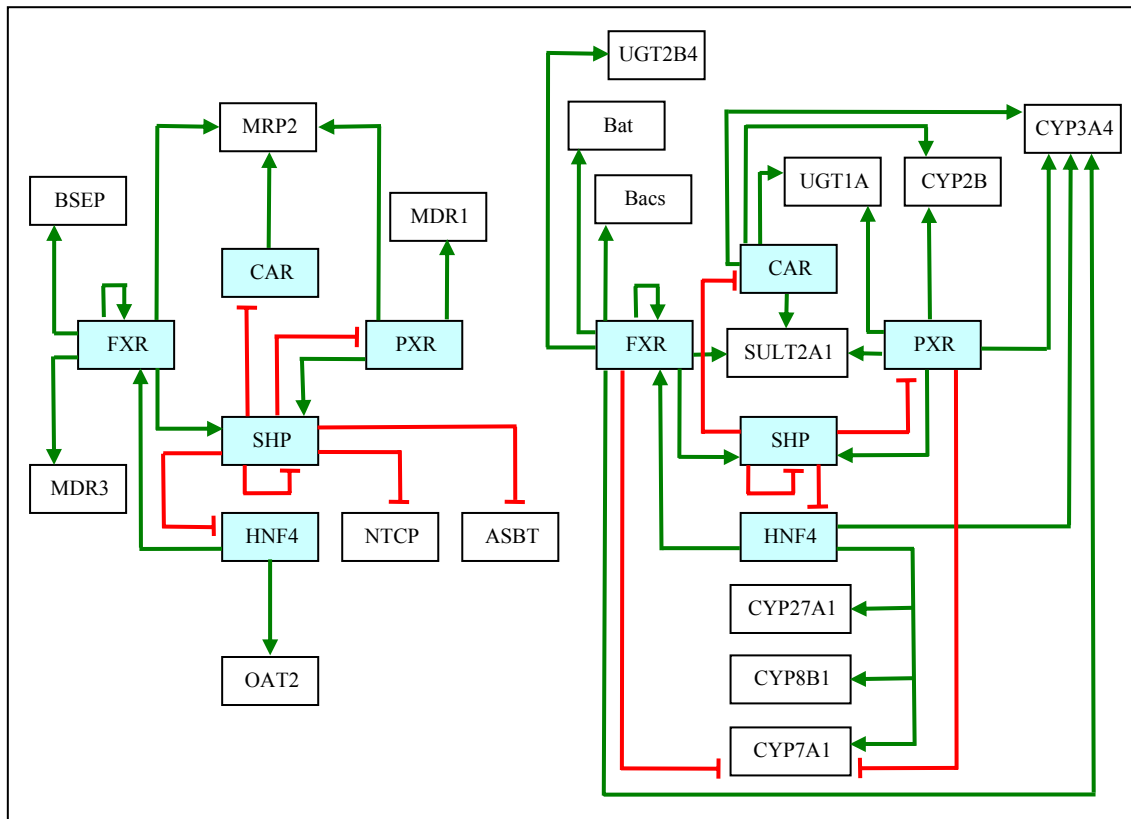


Figure 125: Nuclear receptors FXR, SHP, PXR, CAR and HNF4 regulate a complex network of genes involved in (A) transport and (B) metabolism of drugs and bile acids. Nuclear receptors are shown in light blue boxes. Green and red lines indicate transcriptional activation and suppression. Graph based on Eloranta *et al.*, 2005

In 1999, it was shown that farnesoid X receptor (FXR; NR1H4) a member of the nuclear receptor family was the regulator of bile acid metabolism (Makishima *et al.*, 1999; Parks *et al.*, 1999; Wang *et al.*, 1999). By binding several bile acids and their conjugated derivatives FXR gets activated and stimulates expression of genes involved in bile acid homeostasis like bile salt export pumps (Makishima *et al.*, 1999; Sinal *et al.*, 2000). The rate limiting step of the conversion of cholesterol to bile acids is repressed by FXR triggering a transcriptional cascade of nuclear receptors resulting in a down-regulation of CYP7A1. PXR which is activated by a large set of endogenous and exogenous chemicals is also capable to repress CYP7A1 through a different mechanism. In addition to this key gene of bile acid formation PXR also regulates other genes involved in bile acid metabolism. These genes include multidrug resistance associated protein 2 (ABCC2 MRP2) (Kast *et al.*, 2002), solute carrier organic anion transporter family, member 1B1 (OATP2, SLC21A6) (Staudinger *et al.*, 2001) which

are transporters responsible for bile acid transport. Also regulated by PXR is CYP3A which hydroxylates bile acids including lithocholic acid (LCA) (Araya and Wikvall, 1999). As PXR regulates the previous mentioned genes involved in bile acid metabolism, this receptor next to FXR a second line of defense against the accumulation of toxic bile acids in the liver (Kliwer and Willson, 2002).

In our study induction of primary human hepatocytes with Rifampicin resulted in a decreased expression of cholesterol 7- α hydroxylase (CYP7A) and sterol 12- α hydroxylase (CYP8B1). Bhalla *et al.*, 2004 suggested that ligand activated PXR interferes with HNF-4 signalling by targeting the common co-activator PGC-1. PGC-1 enhances transcriptional activity of HNF-4, a key hepatic activator of CYP7A and CYP8B1. As the HNF-4 based enhancement is repressed by rifampicin-activated PXR the down-regulation of CYP7A and CYP8B1 in Rifampicin treated samples can be explained. Relying on the competition for the co-activator PGC-1 there seems to be a physiologically relevant inhibitory cross-talk between drug- and cholesterol metabolism. In contrast to this observation, CYP3A was strongly induced. The increased CYP3A4 level allows hydroxylation of bile acids and LCA. An interesting idea is to administer patients suffering from cholestasis PXR agonists to use the chemo protective actions of this regulatory circuit. Activated PXR will increase levels of active CYP3A which will hydroxylate bile acids like LCA and form less toxic metabolites (6-hydroxylated bile acids or hyodeoxycholic acid). Interestingly the PXR ligand Rifampicin has been used successfully in the treatment of pruritus caused by cholestasis. In some cases even slight remission of cholestasis was achieved (Bachs *et al.*, 1989; Gillespie and Vickers, 1993; Cancado *et al.*, 1998,). Based on results of the anticholestatic effects of Rifampicin the hope is, that more potent PXR ligands may prove to be even more effective drugs for the treatment of biliary cholestasis.

The whole regulation is even more complex however it can be summarized that drugs and bile acids are imported into the hepatocytes via organic anion transporters (OATs). After hydroxylation and conjugation the more hydrophilic products are excreted into bile or urine. Drug metabolites are transported via ATP-dependent multidrug resistance proteins (MDR1 and MRPs) while bile acids are excreted mainly by bile salt export pump (BSEP; ABCB11). The genes fulfilling the tasks of up-take, metabolism and excretion are regulated by a complex network of transcriptional cascades. Involved in

this regulation are the nuclear receptors FXR and PXR as well as HNF-4. The bile acid synthesizing enzymes CYP7A1 and CYP8B1 are regulated by a negative feedback mechanism based on bile acid concentration which is modulated by co-activators and co-repressors. The role of transcriptional cofactors like PGC-1 mediating and balancing the gene specific effects of individual nuclear receptors is becoming more and more evident and increasingly important (Eloranta and Kullak-Ublick, 2005).

As drug metabolism and hepatic bile salt metabolism are regulated by the same nuclear receptor (PXR) interference between these two functions was expected in the experiments performed in this study. As mentioned Rifampicin induction was thought to interact with three main steps of hepatic bile acid metabolism by altering the expression of transporters and metabolizing enzymes. 30 years ago it was observed in isolated rat hepatocytes that Rifampicin inhibited the hepatic uptake of cholate and taurocholate (Anwer *et al.*, 1978). However not only the uptake of bile acids and other organic anions (bilirubin) is affected, but also the conjugation step of bile acids is changed. Finally the bile salt excretion might be changed as well. To investigate this interference we analysed the expression patterns of genes involved in bile acid metabolism. In Figure 126 the expression of genes from all phases of metabolism are shown: CYP3A4 (Phase I: oxidation), UGT1A1 (Phase II: conjugation) and MRP2 (ABCC2) and MRP3 (ABCC3) (Phase III: transport). The gene expression levels of these four genes are represented as coloured squares. Yellow to red stands for high expression while green to blue represents low expression. Different rows of squares show results from different probe-sets on the array (three for CYP3A4 and three for UGT1A1, one for the transporters).

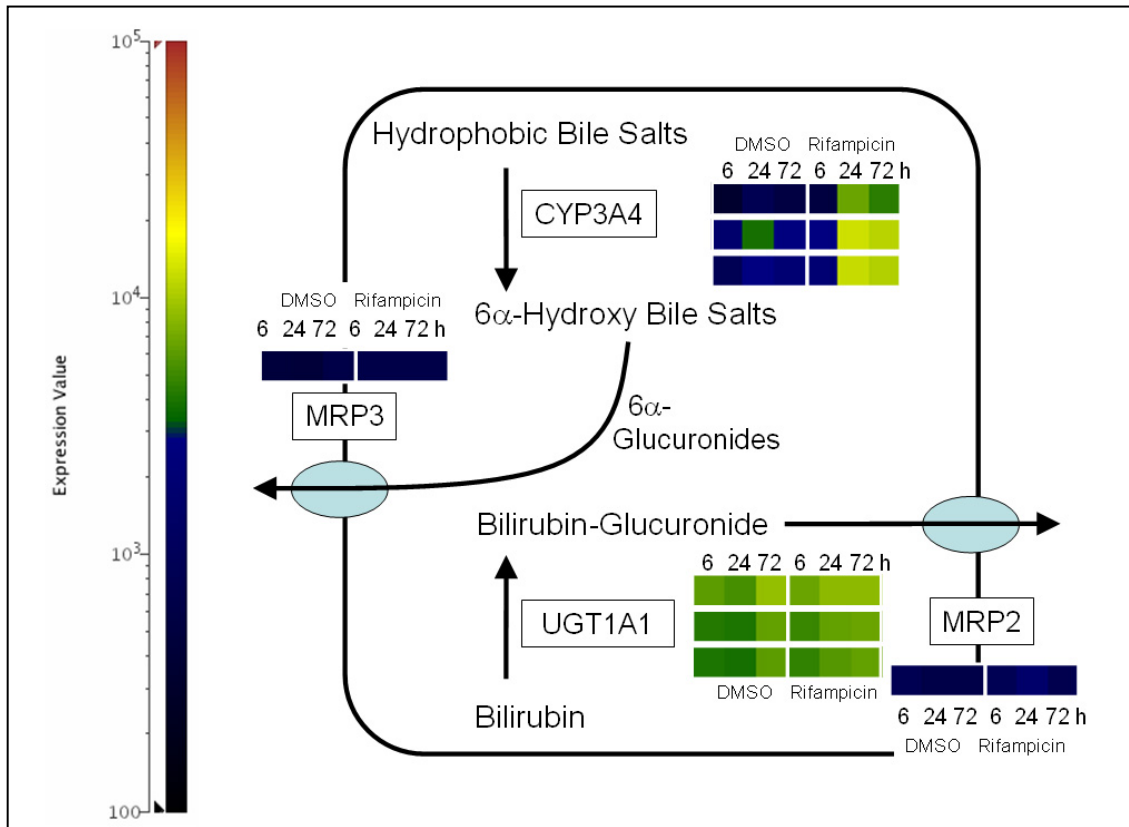


Figure 126: Effects of Rifampicin on hepato-biliary transport system, bile acid and bilirubin metabolizing enzymes. Rifampicin dosage might interfere with bile acid and bilirubin pathways by (1) slightly inducing the bilirubin glucuronyl transferase UGT1A1 and (2) inducing the bilirubin-glucuronide transport protein MRP2. In both cases only a slight up-regulation was detectable. Rifampicin as the prototype inducer of CYP3A4 facilitates conversion of hydrophobic bile acids into hydrophilic 6-hydroxylated compounds which become glucuronidated and most likely are excreted via MRP3. CYP3A4 showed a strong up-regulation while MRP3 was only poorly expressed and showed no induction. Figure based on Gastroenterology 2005 Aug;129(2):476-85.

CYP3A4 catalyzes the reaction from hydrophobic bile salts into hydroxy bile salts which might be also further conjugated with glucuronides and excreted via MRP3 (ABCC3), MRP4 (ABCC4) or organic solute transporter alpha/beta (OST α/β). The transporter MRP2 (ABCC2) mediates excretion of various anions such as bilirubin. UGT1A1 transforms bilirubin into the glucuronide which can be excreted. The hypothesis suggested by Marschall *et al.*, 2005 was that induction with Rifampicin has the following effects on hepato-biliary transport system and bile acid and bilirubin metabolizing enzymes: Rifampicin enhances excretion of bilirubin by inducing the bilirubin glucuronyl transferase UGT1A1 and inducing the bilirubin-glucuronide transport protein ATP-binding cassette, sub-family C (CFTR/MRP), member 2

(ABCC2, MRP2). Rifampicin as the prototype inducer of CYP3A4 facilitates conversion of hydrophobic bile acids into hydrophilic 6-hydroxylated compounds that become glucuronidated and most likely are excreted via ATP-binding cassette, sub-family C (CFTR/MRP), member 3 (ABCC3, MRP3). As shown in Figure 126 the metabolizing enzymes CYP3A4 and UGT1A1 showed an increased expression after stimulation with Rifampicin. In contrast to this observation the expected up-regulation of transporters (MRP2 and MRP3) based on the Rifampicin stimulus was not detected. Only a slight induction by a factor of 1.5-fold was observed for MRP2. The general expression of these transporters was low. A reason for this observation can maybe found in the experimental setup, as the experiments were performed in cellular culture lacking the connection to the bile and other external stimuli. The conclusion of this observation is that for induction of CYP3A4 and UGT1A1 Rifampicin stimulation is sufficient. For induction of gene expression of the involved transporters (MRP2 and MRP3) co-stimulation with other factors is needed.

4.2.3 Glucose- and Fat Metabolism - the Role of Nutritional Regulation

As mentioned previously insulin is the major actor in control of carbohydrate, lipid and protein metabolism restraining processes that release stored energy like lipolysis and ketogenesis, glycogenolysis, proteolysis and gluconeogenesis. As insulin is part of the culture medium a down-regulation of lipolysis and gluconeogenesis was observed in this study. What is even more interesting is the fact that the regulation of gluconeogenesis also interferes with the master regulator of detoxification PXR and its co-activators (<http://www.montp.inserm.fr/u632/>). The connection between gluconeogenesis, fat metabolism and drug metabolism is given by a cross-talk between PXR/CAR and FoxO1. Fox (Forkhead box) genes are a large family of transcription factors controlling a lot of biological functions including gluconeogenesis (Lehmann *et al.*, 2003).

Hepatic gluconeogenesis is tightly controlled and plays an important role for survival during fasting or starvation. In the liver, gluconeogenesis is controlled positively by glucocorticoids, cAMP and glucagon, and negatively by insulin and glucose. Two proteins which play also a regulatory role in xenobiotic metabolism FoxO1 and PGC-1 α

were identified to be involved in regulation of gluconeogenesis (Yoon *et al.*, 2001; Herzig *et al.*, 2001). PGC-1 α induces gluconeogenesis and functions as a co-activator of FoxO1 (Puigserver *et al.*, 2003). In absence of insulin, FoxO1 positively controls the expression of genes involved in gluconeogenesis. As shown in Figure 127 insulin triggers the phosphorylation of FoxO1 by binding to its receptor and activating phosphatidylinositol 3-kinase and Akt kinase (Nakae *et al.*, 2002).

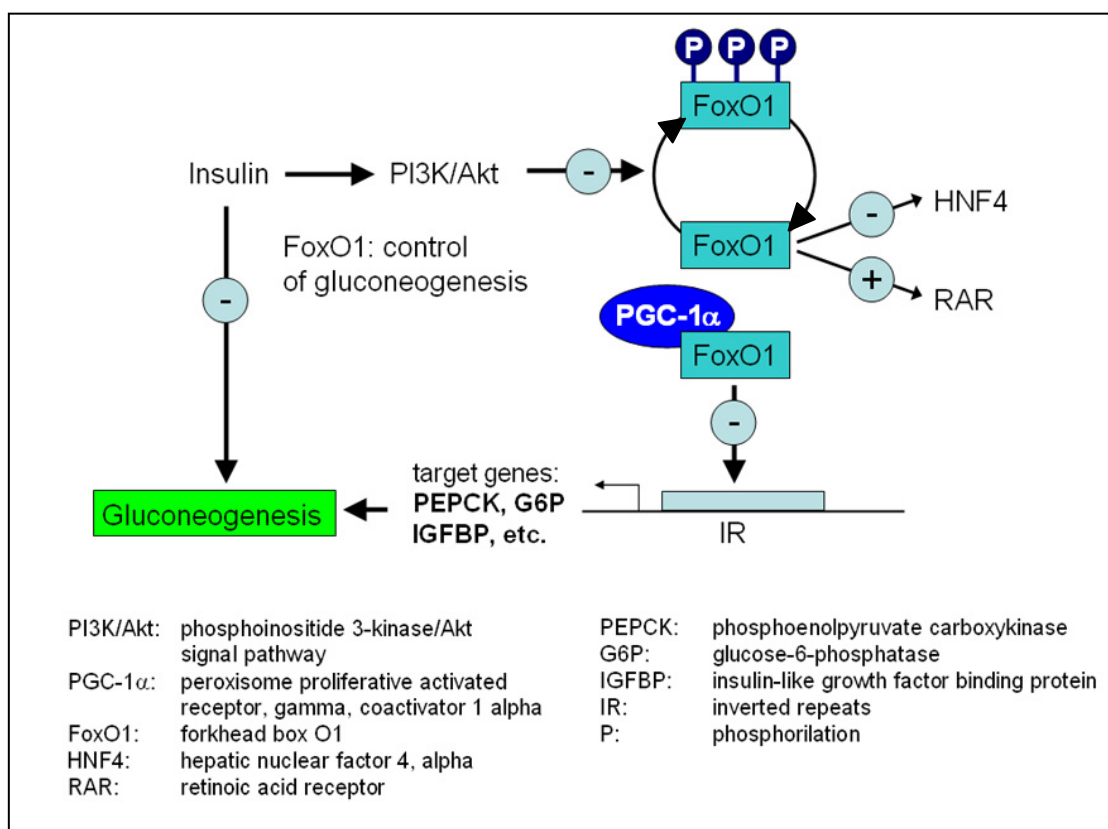


Figure 127: Insulin mediated down-regulation of gluconeogenesis via PI3K/Akt and FoxO1. FoxO1 is the only transcription factor for which a complete path from the insulin receptor to gene regulation has been described. Suppression of gluconeogenesis in hepatocytes might occur based on reduced transcriptional activity/phosphorylation of FoxO1 resulting from the activation of PI3K/Akt signalling cascades. Figure based on Nakae *et al.*, 2002, Zhao *et al.*, 2001, Hirota *et al.*, 2003, Puigserver *et al.*, 2003, Quinn and Yeagley, 2005, Schilling *et al.*, 2006 and Zhang *et al.*, 2006.

Phosphorylated FoxO1 is no longer retained in the nucleus and is unable to bind PGC-1 α which makes it transcriptionally inactive. This is the basis of the negative control of gluconeogenesis by insulin. A recent study also showed that PGC-1 α is unable to induce gluconeogenesis in FoxO1-deficient mouse hepatocytes (Matsumoto *et al.*, 2007). In parallel the cAMP response is significantly reduced. Insulin has no effect on PGC-1 α expression (Puigserver *et al.*, 2003).

FoxO1 has been shown to interact with several nuclear receptors and act as co-repressor (HNF4, GR) or co-activator (RAR) (Hirota *et al.*, 2003; Shimamoto *et al.*, 2004). Kodama and co-workers published the crosstalk between FoxO1 and CAR/PXR (Kodama *et al.*, 2004). Based on a yeast two-hybrid screening, these authors identified FoxO1 as a CAR-binding protein. FoxO1 binds CAR in a ligand-dependent manner and activates its transcriptional activity. In addition FoxO1 similarly binds and activates PXR suggesting that this crosstalk may be of general significance for xenobiotic metabolism. Consistent with these observations, insulin represses the induction of CYP2B10 mRNA in mouse hepatocytes. However the induction of expression of CYP3A4 and CYP2B6 was not or at least only minor affected by insulin in the experiments performed in this work. Maybe another induction mechanism is used, especially as the nuclear receptors were shown to be slightly down-regulated in this study, or the effect of insulin is weaker in primary human hepatocyte cultures.

A strong induction of CYPs indicating an increased drug metabolism in the Rifampicin treated cell cultures compared to controls. Further down-regulation of glycogen synthase 2 (liver, GYS2) and phosphoenolpyruvate carboxykinase 1 (PCK1) show a decrease in gene expression of genes involved in gluconeogenesis. PCK 1 is the enzyme which bypasses together with pyruvate carboxylase the irreversible transfer of a phosphate group from phosphoenolpyruvate (PEP) to ADP, yielding one molecule of pyruvate and one molecule of ATP catalyzed by pyruvate kinase enzyme.

The results based on transcription measurements indicate that drug metabolism and gluconeogenesis are reciprocally co-regulated in response to insulin and xenobiotics (Figure 128). These results are consistent with previous observations made in rat, revealing functional links between insulin- and xenobiotic-mediated regulatory pathways (Sidhu and Omiecinski, 1999). A recent study linked FoxO1 based insulin signalling to c/EBPalpha regulating gluconeogenesis (Sekine *et al.*, 2007). This might be also another connection to the xenobiotic metabolism. The crosstalk between glucose and xenobiotic metabolism might also have clinical implications as diabetes might increase activity of hepatic drug metabolism. Also the other way round might be a useful approach by stimulating drug metabolism to reduce blood glucose levels in diabetic patients.

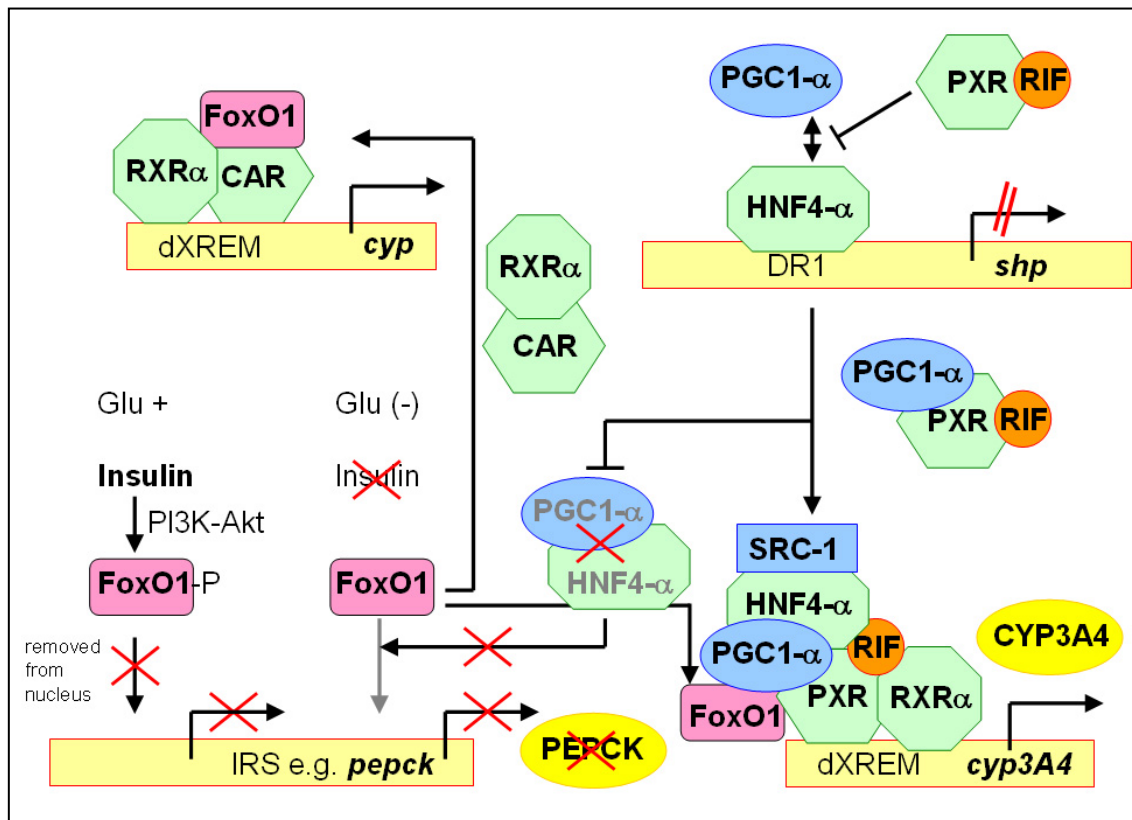


Figure 128: Reciprocally co-regulated drug metabolism and gluconeogenesis based on CAR/RXR-, PXR/RXR FoxO1 cross talk. If glucose is present (Glu +), insulin levels are high, activating PI3K-Akt pathway phosphorylating FoxO1 which is then excluded from the nucleus. This exclusion leads to a down-regulation of gluconeogenetic enzymes like PEPCK. It also interferes with drug metabolism as FoxO1 is essential or at least enhances CYP expression. If glucose levels are low (Glu -) no insulin is present and FoxO1 remains in the nucleus and can activate gluconeogenesis as well as drug metabolism. Drug activated PXR competes with HNF4-α for the co-activator PGC1-α. The dissociation of PGC1-α from the complex with HNF4-α has two effects. First down-regulation of SHP, allowing expression of CYPs and second induction of gluconeogenetic enzymes as the complex HNF4-α/PGC-1α is required for this induction. Drug activated PXR as well as CAR/RXR dimers recruit FoxO1 and form a detoxification complex binding to dXREM leading to an activation of drug metabolism and a down-regulation of gluconeogenesis.

PGC1-α: (PPARGC1A) peroxisome proliferator-activated receptor gamma, coactivator 1 alpha; SRC-1: (NCOA1) p160/steroid receptor coactivator (SRC) family; FoxO1: forkhead box O1; RXRα: retinoid X receptor, alpha; CAR: nuclear receptor subfamily 1, group I, member 3, constitutive androstane receptor; PXR: nuclear receptor subfamily 1, group I, member 2, pregnane X receptor; SHP: (NR0B2) nuclear receptor subfamily 0, group B, member 2, small heterodimer partner; RIF: Rifampicin; Glu: glucose; PEPCK: phosphoenolpyruvate carboxykinase; CYP3A4: cytochrome P450, family 3, subfamily A, polypeptide 4; dXREM: distal xenobiotic responsive element module; DR1: direct repeat element 1; IRS: inverted repeat sequence; PI3K-Akt: phosphoinositide 3-kinase/Akt signal pathway

As discussed by Maurel the physiological objective of this crosstalk might be NADPH homeostasis. NADPH is essential for cytochrome P450-dependent monooxygenase activities and glutathione recycling. In the liver, the pentose phosphate pathway

converts glucose 6-phosphate to ribose 5-phosphate by glucose 6-phosphate dehydrogenase (G6PDH) and generates NADPH. In the gluconeogenesis pathway, glucose 6-phosphate is the last intermediate to glucose. Repression of gluconeogenesis by xenobiotic-activated CAR/PXR might be required to maintain sufficient NADPH levels for xenobiotic metabolism. Interestingly, patients deficient in G6PDH exhibit reduced detoxification ability (Efferth *et al.*, 2006). However further experiments are needed to validate this observation.

The ability to alter the sensitivity of a key transcription factor to improve Insulin regulated control of blood glucose would be a major improvement in the treatment of diabetes (Quinn and Yeagley, 2005)

4.2.4 Induction of Metallothionein Proteins

Metallothioneins (MTs) are members of a family of conserved, ubiquitous, cysteine rich, low molecular weight (MW ranging from 3.5 to 14 kDa) proteins. MTs have the capacity to bind both physiological (Zn, Cu, Se) and xenobiotic (Cd, Hg, Ag) heavy metals through the thiol group of its cysteine residues, which represents almost 30% of its amino acidic residues.

MT are implicated in a diversity of intracellular functions (Davis and Cousins, 2000), but the only consensus among researchers so far concerns its role in the detoxification of heavy metals, which is due mostly to its high affinity for these metals (Klassen *et al.*, 1999; Templeton and Cherian, 1991). Data published suggests MTs to be involved in protection against metal toxicity, regulation of physiological metals (Zn and Cu) and protection against oxidative and alcoholic stress (Formigari *et al.*, 2007). There are four main isoforms expressed in humans. In the human body, large quantities are synthesised primarily in the liver and kidneys.

Several members of metallothionein proteins (MT 1F, 1X, 1E, 1M, 1G, 2A and 1H) were found to be significantly up-regulated in Rifampicin treated cultures compared to control (DMSO) cultures. Metallothionein gene expression is induced by a high variety of stimuli, as metal exposure, oxidative stress, radiation (Roudkenar *et al.*, 2007), hydric stress and glucocorticoids. In mouse liver the up-regulation of metallothionein proteins was implicated with reduced alcoholic liver injury caused by Maotai (Liu *et*

al., 2006). Maotai is a commonly consumed beverage in China containing 53 % alcohol. In contrast to ethanol in equal amounts it is less toxic and induces metallothionein. Therefore the authors suggested a beneficial effect of the induction of MTs for Maotai/ethanol detoxification and reduced liver toxicity.

The level of the response to these inducers depends on the MT gene. MT genes present in their promoter specific sequences for the regulation of expression like metal response elements (MRE), glucocorticoid response elements (GRE) or antioxidant response elements (ARE). An overview of possible regulation mechanisms and functional roles are summarized in Figure 129.

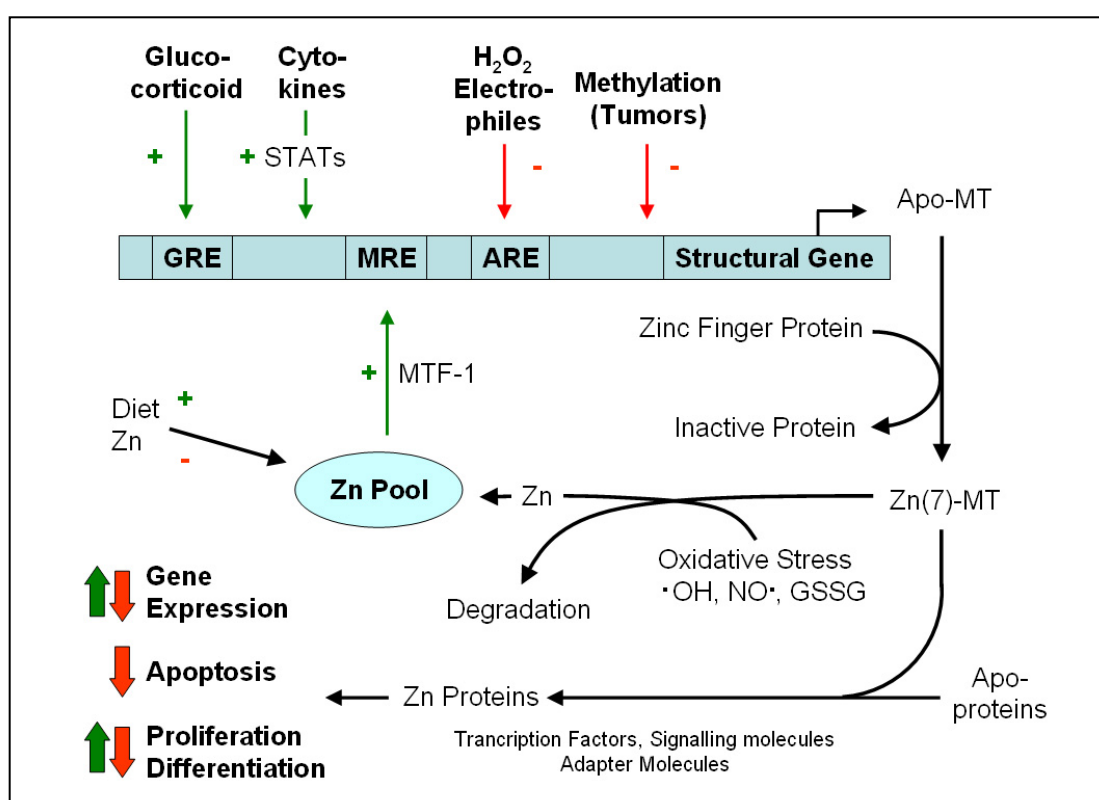


Figure 129: Overview of metallothionein (MT) gene regulation and function. MT genes present in their promoters specific sequences for the regulation of expression. Figure based on Davis and Cousins, 2000.

Binding of zinc activates metal responsive transcription factor (MTF-1) which binds to the metal response element (MRE). Glucocorticoids can also activate gene transcription via glucocorticoid response elements (GRE). Cytokine signals are promoted by signal transducers and activators of transcription (STAT). Redox status triggers expression via antioxidant response element (ARE). In tumour cells methylation might down-regulated

the expression of MT proteins. Intracellular zinc pools are balanced by dietary zinc uptake. This pool is the source for zinc incorporated into MT. Apo MT (thionein) take up zinc while Zn(7)-MT may donate zinc. The numerous zinc coordination sites of proteins like transcription factors, signalling molecules and adapters which use zinc fingers for protein-protein interactions provide the opportunity for MT by altering Zn availability to influence cellular key processes. The response to oxidative damage caused by oxidative stress and electrophiles could be regulated as well as basic functions like gene regulation, cell proliferation, differentiation, signal transduction or apoptosis.

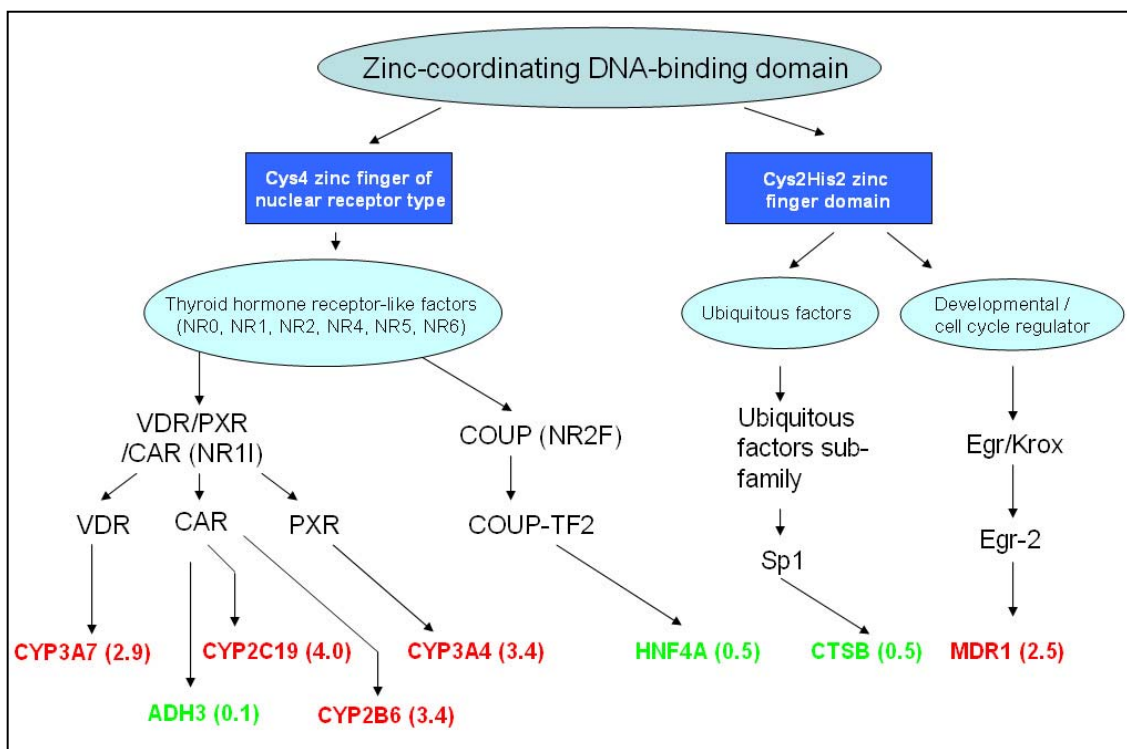


Figure 130: Potential role of zinc in regulation of gene expression. Shown is the connection between zinc-coordinating DNA-binding domains of gene regulatory proteins like nuclear receptors (NR) and zinc finger domain proteins. These proteins are linked to some of their target genes. The number indicates the state of expression, red means up-regulated, green means down-regulated into human hepatocytes after 24 h of Rifampicin induction.

One example from this study is carbonic anhydrase 12 which was shown to be slightly up-regulated in Rifampicin treated cells. It is one member of a large family of zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide ($\text{CO}_2 + \text{H}_2\text{O} \leftrightarrow$

H₂CO₃) and participates in a variety of biological processes, including respiration, calcification, acid-base balance, bone resorption, and the formation of aqueous humor, cerebrospinal fluid, saliva, and gastric acid. The expression of this zinc dependent gene might be altered by the metallothionein based regulation.

The role in zinc circulation has broad influence of cellular responses mediated by MTs. In the experiments performed in this work an increase of MT mRNA levels was observed maybe based on stress, especially oxidative stress co-occurring with increased monooxygenase metabolism. Also modulating differentiation mechanisms of the hepatocytes or transcription in general in a zinc dependent manner, as shown in Figure 130 may require an up-regulation of MTs.

4.2.5 Down-regulation of Cadherin Cell Adhesion Receptor

In this study a strong down-regulation of Cadherin 1 (CDH1) over time was observed. From a strong induced state (3.5-fold over-expression in Rifampicin induced cells) after 6 h the expression was going down to 2-fold down-regulation after 72 h in Rifampicin cell cultures compared to control cell cultures (DMSO).

Cadherins are transmembrane proteins which play important roles in cell adhesion, ensuring that cells within tissues are bound together. They are dependent on calcium ions to function, hence their name „Calcium adhering“ (Ca-adherine). Cadherins are cell-cell adhesion receptors which are expressed in all cell types that form solid tissues. The regulated expression of cadherins controls cell polarity, tissue morphology and homotypic cell aggregation (Takeichi, 1991). Loss of cadherins induces dispersion of cells. While cadherins have been classically studied for their role as mediators of cell segregation and tissue morphogenesis (Takeichi, 1988), additional interest in cadherins was generated by reports that they may play a role in cyto-differentiation (Knudsen *et al.*, 1998) and are associated with cell-cell contact polarity maintaining differentiated functions in cultured hepatocytes (Moghe *et al.*, 1996). The observed down-regulation of cadherin is an interesting result as Brieva and Moghe, 2001 showed, that functional activity of liver cell cultures requires the modulation of cell-cell interactions based on varied presentation of epithelial- (E-)cadherin. Hepatocytes co-cultured with cadherin-presenting chaperone cells had a 55 -56 % increase in longterm function (Brieva and

Moghe, 2001). The induced increase in hepatocyte biochemical functions (albumin and urea secretion) were governed by the degree of cadherin based contacts presented during culture. A recent report showed that in cryopreserved hepatocytes degradation of E-cadherin was increased compared to fresh hepatocytes (Terry *et al.*, 2007). Cell attachment to collagen-coated plates was monitored and in parallel the expression of adhesion related molecules was measured. Attachment efficiency was significantly decreased after cryopreservation. As good quality cryopreserved human hepatocytes are becoming an important source for clinical applications new cryopreservation protocols are needed. Furthermore additional experiments are needed to shed further light on the role of cadherins in cell culture experiments in general and maybe also in Rifampicin induction.

4.2.6 Effects of Inflammatory Cytokines on Expression of Cytochrome P450 Enzymes

The activity of cytochrome P450 (CYP) changes during disease, injury or in general when inflammatory response is activated (Renton, 2005). Cytokines released during inflammatory response are found to be responsible for CYP repression. During time series experiments with sub-genome arrays an inhibition of CYP3A4 was observed in this study. After strong down-regulation of IL-6 in cell culture CYP3A4 was finally induced. Therefore a negative regulation of CYP3A4 by IL-6 the main regulator of the hepatic acute phase response is supported. Most of the major CYP forms are affected by inflammation leading to a decrease in the capacity of the liver to handle drugs, chemicals and other endogenous compounds. The loss in metabolism capacity influencing the fate and therapeutic efficacy of many drugs seems to be an effect resulting from altered interaction of transcription factors that control the expression of CYPs and other enzymes. Inflammatory response starts with production of interleukin 1 β (IL-1 β), tumour necrosis factor α (TNF- α) and interleukin 6 (IL-6) which are released into the blood stream. This initiating step is the so called acute phase response (APR) (Baumann and Gauldie, 1994; Koj, 1996). Hepatocytes are influenced by cytokines and express acute phase proteins (APPs) (Castell *et al.*, 1990). IL-6 is recognized as being the most important cytokine in the hepatic response during

inflammation and the major regulator of hepatic APP synthesis (Castell *et al.*, 1989; Heinrich *et al.*, 1990). CYP3A4 down-regulation does not occur via Janus kinases (JAK) and signal transducers and activators of transcription (STAT) pathways. Also the involvement of IL-6 activated kinases like ERK1/2 or p38 is unlikely (Jover *et al.*, 2002). However the glycoprotein receptor gp130 is required as common receptor component for IL-6 like cytokines. The binding of the cytokine to the receptor activates cytoplasmatic signal cascades (Heinrich *et al.*, 1998; Taga and Kishimoto, 1997). Cytokines can act by altering the availability of transcription factors like CCAAT-enhancer binding proteins (c/EBPs). Expression of different c/EBPs is modulated in response to inflammatory cytokines (Alam *et al.*, 1992). For example a slight induction of transcription factor c/EBP β -LIP (liver enriched transcriptional inhibitory protein) is caused by IL-6. As c/EBP β -LIP has no transactivation domain it causes a repression of CYP3A4 expression. A possible mechanism explaining the down-regulation of CYP is that IL-6 down-regulates CYPs through translational induction of c/EBP β -Lip competing with constitutive c/EBP transactivators (Figure 131).

The expression change of detoxifying enzymes during the APR seems to be based on the balance between the concentrations of the two isoforms of c/EBP β mRNA, LAP (liver enriched transcriptional activating protein) and LIP. While over-expression of c/EBP α and c/EBP β -LAP causes an up-regulation in human hepatoma cells, the LIP-form down regulates CYP3A4 expression by antagonizing the activation effects of c/EBP α and c/EBP β -LAP (Jover *et al.*, 2002). From a clinical point of view this observation has strong impact on drug prescription. For any drug that is metabolized by CYPs used during infection or inflammatory response there is a significant risk leading to an adverse drug response, especially when this drugs have a narrow therapeutic index. For development of chemotherapeutics this fact has taken into consideration as there is a strong link between cancer associated inflammation and impaired drug metabolism (Charles *et al.*, 2006). Inflammation is a common damage associated response. Also in therapeutic medication down-regulation of CYPs is an important issue. The increasing number of therapeutic cytokines, like interferons, IL-2, IL-11, erythropoietin or granulocyte-colony stimulating factor might have a repressive potential on drug metabolism. To overcome these side effects therapeutical cytokines

have to be developed showing no or a smaller effect on hepatic drug metabolizing enzymes.

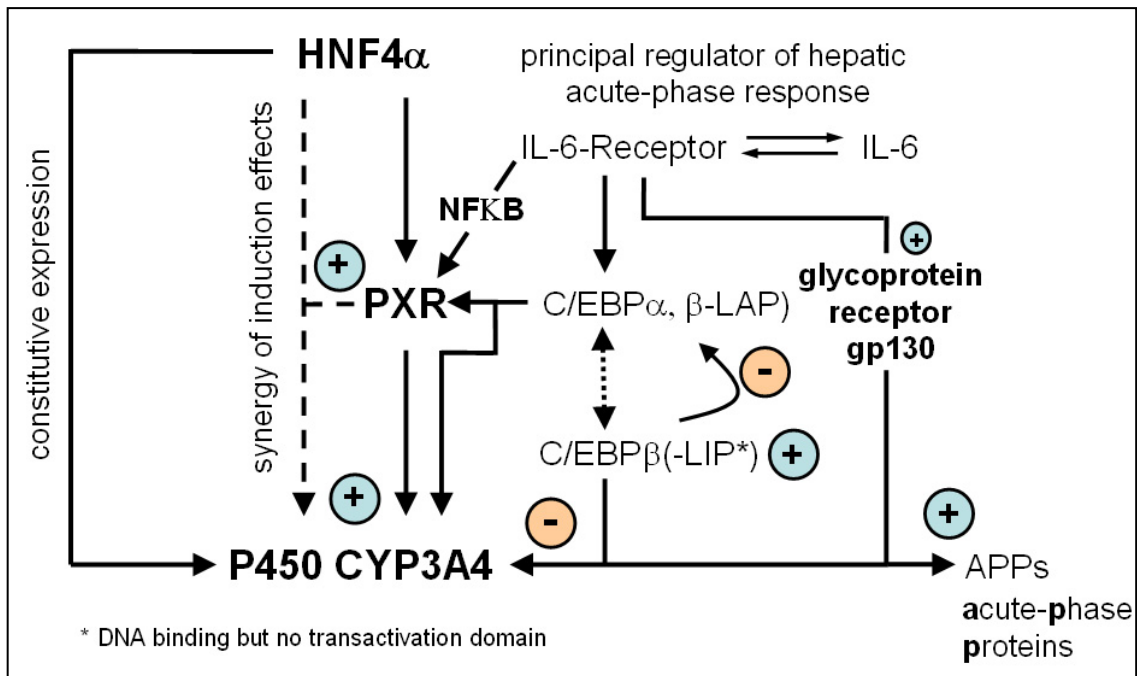


Figure 131: Suggested interaction of inflammation with parts of the network coordinating regulation of P450 CYP3A4 via PXR, HNF4α and other factors. The negative regulation of CYP3A4 by interleukin 6 (IL-6) requires activation of the glycoprotein receptor gp130. CYP3A4 might be down-regulated by IL-6 through translational induction of C/EBPβ-LIP. C/EBPβ-LIP: DNA binding but no transactivation domain, LAP: liver-enriched transcriptional activating protein, LIP: liver-enriched transcriptional inhibitory protein. Circles indicate stimulation (blue "+") or repression (red "-") of function. This overview was derived from expression data obtained from the work presented here and literature (Jover *et al.*, 2002).

4.3 Conclusion

The presented thesis work provides an overview of different DNA-microarray technology platforms from one colour short oligo gene chips (Affymetrix) over two colour long oligo chips (Agilent) to one- and two-colour cDNA (X-mer) arrays from Eppendorf. The platform from Affymetrix is based on indirect labeling while Agilent uses a direct labeling protocol. Eppendorf arrays allow both direct- and indirect labeling. The comparison of these three array platforms from Affymetrix, Agilent and Eppendorf showed an overall poor correlation concerning quantification of expression

results. However in a Boolean framework of 0 and 1 meaning gene up-regulated or down-regulated a very satisfactory result could be obtained using these different platforms. Several hints on how to analyse these huge data sets obtained from gene expression experiments using DNA microarrays are given.

Detailed plans on working with human cell cultures for time dependant perturbation experiments are given. Furthermore three different RNA isolation kits were extensively compared due to yield, reproducibility, purity, integrity, cost and handling.

This technical part of the thesis gives other investigators helpful suggestions on how to plan and perform array experiments and to avoid common pitfalls.

The microarray screening of gene expression was applied to the analysis of the time dependent response of primary human hepatocyte cells to a stimulus with the antibiotic Rifampicin. At first several already known key player of the detoxification process could be validated being up-regulated. The genes found to be differentially expressed like CYP3A4, CYP2B6, ABCB1, ALAS1 that confirm published knowledge indicate reliable performance of the experiments and proof that the used samples showed a normal, common behaviour. Furthermore several observations fitted well into the expectations but were not described in detail so far. Examples are the interference of drug metabolism, bile acid metabolism and fatty acid homeostasis as well as carbon metabolism. Furthermore the interaction of inflammation (induction of IL-6) with drug metabolism (repression of CYP3A4) based on interleukins was shown on mRNA level. Other observations like the down-regulation of nuclear receptors (NR1I2 and NR1I3) as well as HNF4 α were in contrast to previous reports from experiments in animal models. Further independent confirmation of gene expression change for more genes has to be performed e.g. based on real-time PCR measurements followed by investigations on protein level (Western-Blot and immunohistochemistry) and metabolite studies validating the obtained results. In addition silencing of genes coding for regulatory proteins or candidates would nowadays also be an interesting approach to elucidate functions. However all subsequent analysis rely on high throughput full genome identification of suitable targets. At the moment the only technology which can provide this requirement is the microarray technology used in the study presented here. Based on the observations and collected expression data, modelling approaches are ongoing to translate the data into mathematical equations which map the biological response on

gene expression to a computer model. These models will allow prediction of cellular behaviour. It will be possible to deduce new interactions and interaction partners of e.g. regulatory circuits as shown in this work in collaboration with P. K. Murugan (IBVT).

A similar approach like it is used in biology nowadays and was applied in this work to describe a complex system can be found 400 years ago in the field of astronomy. Johannes Kepler investigated basic principles of a complex system, describing the movement of planets. Finally Isaac Newton was the scientist who developed based on Kepler's laws of planetary motion the theory of gravitation, thus removing the last doubts about heliocentrism and advancing the scientific revolution.

In the study presented here the change of gene expression was monitored and also basic rules of induction and interaction were described. Now it is the time in systems biology for people building models using this rich collection of data for further investigations of complex system like the hepatocyte. Their task is to come up with rules and laws describing nature, launching a scientific revolution in the post genome area. First simulation and then prediction of the behaviour of regulatory and metabolizing circuits is the goal of this task. Finally the evolving models unravelling basic principles of signal transduction in life, will allow forecasts of the properties and dynamic behaviour of biological systems. Especially for drug companies with the philosophy "fail early - fail cheaply" this predictive power will reduce costs and facilitate new drug-development. This is not only because model based prediction will allow identification of new targets but also make a pre-selection of the most promising ones and to forecast adverse effects. The *in-silico* hepatocyte as one example for a model evolved in a systems biology approach will allow a better personalized medicine which will be more reliable, affordable and it will also have strong impact on (medical-) research in general.

5 Literature

- Acocella, G. (1978). Clinical pharmacokinetics of rifampicin. *Clin Pharmacokinet* 3, 108-127.
- Adams, R., and Bischof, L. (1994). Seeded region growing. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE Vol. 16*, 641 - 647.
- Alam, T., An, M.R., and Papaconstantinou, J. (1992). Differential expression of three C/EBP isoforms in multiple tissues during the acute phase response. *J Biol Chem* 267, 5021-5024.
- Anwer, M.S., Kroker, R., and Hegner, D. (1978). Inhibition of hepatic uptake of bile acids by rifamycins. *Naunyn Schmiedebergs Arch Pharmacol* 302, 19-24.
- Araya, Z., and Wikvall, K. (1999). 6alpha-hydroxylation of taurochenodeoxycholic acid and lithocholic acid by CYP3A4 in human liver microsomes. *Biochim Biophys Acta* 1438, 47-54.
- Bachs, L., Pares, A., Elena, M., Piera, C., and Rodes, J. (1989). Comparison of rifampicin with phenobarbitone for treatment of pruritus in biliary cirrhosis. *Lancet* 1, 574-576.
- Ball, C., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Icahn, C., Parkinson, H., Quackenbush, J., Ringwald, M., Sansone, S.A., Sherlock, G., Spellman, P., Stoeckert, C., Tatenos, Y., Taylor, R., White, J., and Winegarden, N. (2004). An open letter on microarray data from the MGED Society. *Microbiology* 150, 3522-3524.
- Ball, C.A., and Brazma, A. (2006). MGED standards: work in progress. *Omics* 10, 138-144.
- Bammler, T., Beyer, R.P., Bhattacharya, S., Boorman, G.A., Boyles, A., Bradford, B.U., Bumgarner, R.E., Bushel, P.R., Chaturvedi, K., Choi, D., Cunningham, M.L., Deng, S., Dressman, H.K., Fannin, R.D., Farin, F.M., Freedman, J.H., Fry, R.C., Harper, A., Humble, M.C., Hurban, P., Kavanagh, T.J., Kaufmann, W.K., Kerr, K.F., Jing, L., Lapidus, J.A., Lasarev, M.R., Li, J., Li, Y.J., Lobenhofer, E.K., Lu, X., Malek, R.L., Milton, S., Nagalla, S.R., O'Malley J, P., Palmer, V.S., Pattee, P., Paules, R.S., Perou, C.M., Phillips, K., Qin, L.X., Qiu, Y., Quigley, S.D., Rodland, M., Rusyn, I., Samson, L.D., Schwartz, D.A., Shi, Y., Shin, J.L., Sieber, S.O., Slifer, S., Speer, M.C., Spencer, P.S., Sproles, D.I., Swenberg, J.A., Suk, W.A., Sullivan, R.C., Tian, R., Tennant, R.W., Todd, S.A., Tucker, C.J., Van Houten, B., Weis, B.K., Xuan, S., and Zarbl, H. (2005). Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2, 351-356.

Barczak, A., Rodriguez, M.W., Hanspers, K., Koth, L.L., Tai, Y.C., Bolstad, B.M., Speed, T.P., and Erle, D.J. (2003). Spotted long oligonucleotide arrays for human gene expression analysis. *Genome Res* 13, 1775-1785.

Barthel, A., Schmolli, D., and Unterman, T.G. (2005). FoxO proteins in insulin action and metabolism. *Trends Endocrinol Metab* 16, 183-189.

Baugh, L.R., Hill, A.A., Brown, E.L., and Hunter, C.P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* 29, E29.

Baumann, H., and Gauldie, J. (1994). The acute phase response. *Immunol Today* 15, 74-80.

Benton, W.D., and Davis, R.W. (1977). Screening lambda_{gt} recombinant clones by hybridization to single plaques in situ. *Science* 196, 180-182.

Beucher, S., and Meyer, F. (1993). *The Morphological Approach to Segmentation: The Watershed Transformation*. Marcel Dekker, NY.

Bhalla, S., Ozalp, C., Fang, S., Xiang, L., and Kemper, J.K. (2004). Ligand-activated pregnane X receptor interferes with HNF-4 signaling by targeting a common coactivator PGC-1 α . Functional implications in hepatic cholesterol and glucose metabolism. *J Biol Chem* 279, 45139-45147.

Bogaards, J.J., Bertrand, M., Jackson, P., Oudshoorn, M.J., Weaver, R.J., van Bladeren, P.J., and Walther, B. (2000). Determining the best animal model for human cytochrome P450 activities: a comparison of mouse, rat, rabbit, dog, micropig, monkey and man. *Xenobiotica* 30, 1131-1152.

Boureaux, A., Vignal, E., Faure, S., and Fort, P. (2007). Evolution of the Rho family of ras-like GTPases in eukaryotes. *Mol Biol Evol* 24, 203-216.

Brieva, T.A., and Moghe, P.V. (2001). Functional engineering of hepatocytes via heterocellular presentation of a homoadhesive molecule, E-cadherin. *Biotechnol Bioeng* 76, 295-302.

Brownstein, M. (2006). Sample labeling: an overview. *Methods Enzymol* 410, 222-237.

Bustelo, X.R., Sauzeau, V., and Berenjano, I.M. (2007). GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. *Bioessays* 29, 356-370.

Calleja, C., Eeckhoutte, C., Dacasto, M., Larrieu, G., Dupuy, J., Pineau, T., and Galtier, P. (1998). Comparative effects of cytokines on constitutive and inducible expression of the gene encoding for the cytochrome P450 3A6 isoenzyme in cultured rabbit hepatocytes: consequences on progesterone 6 β -hydroxylation. *Biochem Pharmacol* 56, 1279-1285.

Cancado, E.L., Leitao, R.M., Carrilho, F.J., and Laudanna, A.A. (1998). Unexpected clinical remission of cholestasis after rifampicin therapy in patients with normal or

slightly increased levels of gamma-glutamyl transpeptidase. *Am J Gastroenterol* 93, 1510-1517.

Castell, J.V., Gomez-Lechon, M.J., David, M., Andus, T., Geiger, T., Trullenque, R., Fabra, R., and Heinrich, P.C. (1989). Interleukin-6 is the major regulator of acute phase protein synthesis in adult human hepatocytes. *FEBS Lett* 242, 237-239.

Castell, J.V., Gomez-Lechon, M.J., David, M., Fabra, R., Trullenque, R., and Heinrich, P.C. (1990). Acute-phase response of human hepatocytes: regulation of acute-phase protein synthesis by interleukin-6. *Hepatology* 12, 1179-1186.

Causton, H.C., and Game, L. (2003). MGED comes of age. *Genome Biol* 4, 351.

Charles, K.A., Rivory, L.P., Brown, S.L., Liddle, C., Clarke, S.J., and Robertson, G.R. (2006). Transcriptional repression of hepatic cytochrome P450 3A4 gene in the presence of cancer. *Clin Cancer Res* 12, 7492-7497.

Chen, W., Owsley, E., Yang, Y., Stroup, D., and Chiang, J.Y. (2001). Nuclear receptor-mediated repression of human cholesterol 7 α -hydroxylase gene transcription by bile acids. *J Lipid Res* 42, 1402-1412.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 32 *Suppl*, 490-495.

Cohen, S.N., Chang, A.C., and Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci U S A* 69, 2110-2114.

Comander, J., Natarajan, S., Gimbrone, M.A., Jr., and Garcia-Cardena, G. (2004). Improving the statistical detection of regulated genes from microarray data using intensity-based variance estimation. *BMC Genomics* 5, 17.

Dalma-Weiszhausz, D.D., Warrington, J., Tanimoto, E.Y., and Miyada, C.G. (2006). The affymetrix GeneChip platform: an overview. *Methods Enzymol* 410, 3-28.

Davis, S.R., and Cousins, R.J. (2000). Metallothionein expression in animals: a physiological perspective on function. *J Nutr* 130, 1085-1088.

de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9, 67-103.

de Longueville, F., Atienzar, F.A., Marcq, L., Dufrane, S., Evrard, S., Wouters, L., Leroux, F., Bertholet, V., Gerin, B., Whomsley, R., Arnould, T., Remacle, J., and Canning, M. (2003). Use of a low-density microarray for studying gene expression patterns induced by hepatotoxicants on primary cultures of rat hepatocytes. *Toxicol Sci* 75, 378-392.

deLongueville, F., and Heim, A. (2004). Xmer technology: Probe selection for DualChip Microarrays. Technical Application Note.

- Dennise, D., Dalma-Weiszhausz, D.D., Warrington, J., Tanimoto, E.Y., and Miyada, C.G. (2006). The affymetrix GeneChip platform: an overview. *Methods Enzymol* *410*, 3-28.
- DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* *278*, 680-686.
- DeWitt, D.L., and Smith, W.L. (1988). Primary structure of prostaglandin G/H synthase from sheep vesicular gland determined from the complementary DNA sequence. *Proc Natl Acad Sci U S A* *85*, 1412-1416.
- Dudoit, S., Yang, Y., Callow, M., and Speed, T. (2000). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* *12*, 111 - 139.
- Dussault, I., and Forman, B.M. (2002). The nuclear receptor PXR: a master regulator of "homeland" defense. *Crit Rev Eukaryot Gene Expr* *12*, 53-64.
- Efferth, T., Schwarzl, S.M., Smith, J., and Osieka, R. (2006). Role of glucose-6-phosphate dehydrogenase for oxidative stress and apoptosis. *Cell Death Differ* *13*, 527-528; author reply 529-530.
- Efron, B., Tibshirani, R., Goss, V., and Chu, G. (2000). Microarrays and their use in a comparative experiment. Technical report, Stanford University.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* *95*, 14863-14868.
- Eloranta, J.J., and Kullak-Ublick, G.A. (2005). Coordinate transcriptional regulation of bile acid homeostasis and drug metabolism. *Arch Biochem Biophys* *433*, 397-412.
- Eloranta, J.J., Meier, P.J., and Kullak-Ublick, G.A. (2005). Coordinate transcriptional regulation of transport and metabolism. *Methods Enzymol* *400*, 511-530.
- Falkner, K.C., Ritter, J.K., and Prough, R.A. (2007). Regulation of the Rat Udp-Glycosyltransferase 1a6 by Glucocorticoids Involves a Cryptic Glucocorticoid Response Element. *Drug Metab Dispos*.
- Ferguson, S.S., Chen, Y., LeCluyse, E.L., Negishi, M., and Goldstein, J.A. (2005). Human CYP2C8 is transcriptionally regulated by the nuclear receptors constitutive androstane receptor, pregnane X receptor, glucocorticoid receptor, and hepatic nuclear factor 4alpha. *Mol Pharmacol* *68*, 747-757.
- Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* *251*, 767-773.
- Formigari, A., Irato, P., and Santon, A. (2007). Zinc, antioxidant systems and metallothionein in metal mediated-apoptosis: biochemical and cytochemical aspects. *Comp Biochem Physiol C Toxicol Pharmacol* *146*, 443-459.

- Fromm, M.F., Busse, D., Kroemer, H.K., and Eichelbaum, M. (1996). Differential induction of prehepatic and hepatic metabolism of verapamil by rifampin. *Hepatology* 24, 796-801.
- Gao, X., LeProust, E., Zhang, H., Srivannavit, O., Gulari, E., Yu, P., Nishiguchi, C., Xiang, Q., and Zhou, X. (2001). A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Res* 29, 4744-4750.
- Gillespie, D.A., and Vickers, C.R. (1993). Pruritus and cholestasis: therapeutic options. *J Gastroenterol Hepatol* 8, 168-173.
- Gollub, J., and Sherlock, G. (2006). Clustering microarray data. *Methods Enzymol* 411, 194-213.
- Grunstein, M., and Hogness, D.S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci U S A* 72, 3961-3965.
- Guengerich, F.P. (1997). Comparisons of catalytic selectivity of cytochrome P450 subfamily enzymes from different species. *Chem Biol Interact* 106, 161-182.
- Guo, D., Sarkar, J., Ahmed, M.R., Viswakarma, N., Jia, Y., Yu, S., Sambasiva Rao, M., and Reddy, J.K. (2006). Peroxisome proliferator-activated receptor (PPAR)-binding protein (PBP) but not PPAR-interacting protein (PRIP) is required for nuclear translocation of constitutive androstane receptor in mouse liver. *Biochem Biophys Res Commun* 347, 485-495.
- Handschin, C., Lin, J., Rhee, J., Peyer, A.K., Chin, S., Wu, P.H., Meyer, U.A., and Spiegelman, B.M. (2005). Nutritional regulation of hepatic heme biosynthesis and porphyria through PGC-1alpha. *Cell* 122, 505-515.
- Heinrich, P.C., Behrmann, I., Muller-Newen, G., Schaper, F., and Graeve, L. (1998). Interleukin-6-type cytokine signalling through the gp130/Jak/STAT pathway. *Biochem J* 334 (Pt 2), 297-314.
- Heinrich, P.C., Castell, J.V., and Andus, T. (1990). Interleukin-6 and the acute phase response. *Biochem J* 265, 621-636.
- Herzig, S., Long, F., Jhala, U.S., Hedrick, S., Quinn, R., Bauer, A., Rudolph, D., Schutz, G., Yoon, C., Puigserver, P., Spiegelman, B., and Montminy, M. (2001). CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature* 413, 179-183.
- Hirota, K., Daitoku, H., Matsuzaki, H., Araya, N., Yamagata, K., Asada, S., Sugaya, T., and Fukamizu, A. (2003). Hepatocyte nuclear factor-4 is a novel downstream target of insulin via FKHR as a signal-regulated transcriptional inhibitor. *J Biol Chem* 278, 13056-13060.
- Hodgson, E., and Rose, R.L. (2007). Human metabolic interactions of environmental chemicals. *J Biochem Mol Toxicol* 21, 182-186.

Holland, P.M., Abramson, R.D., Watson, R., and Gelfand, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A* 88, 7276-7280.

Honkakoski, P., Sueyoshi, T., and Negishi, M. (2003). Drug-activated nuclear receptors CAR and PXR. *Ann Med* 35, 172-182.

Ideker, T. (2004). Systems biology 101--what you need to know. *Nat Biotechnol* 22, 473-475.

Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A., and Auffray, C. (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* 33, e56.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31, e15.

Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S.C., Hoffman, E., Jedlicka, A.E., Kawasaki, E., Martinez-Murillo, F., Morsberger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A., Wilson, M., Yang, Y., Ye, S.Q., and Yu, W. (2005). Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2, 345-350.

Jarvinen, A.K., Hautaniemi, S., Edgren, H., Auvinen, P., Saarela, J., Kallioniemi, O.P., and Monni, O. (2004). Are data from different gene expression microarray platforms comparable? *Genomics* 83, 1164-1168.

Jayapal, M., and Melendez, A.J. (2006). DNA microarray technology for target identification and validation. *Clin Exp Pharmacol Physiol* 33, 496-503.

Johnson, S.R., and Pavord, I.D. (1996). Grand Rounds--City Hospital, Nottingham. A complicated case of community acquired pneumonia. *Bmj* 312, 899-901.

Jover, R., Bort, R., Gomez-Lechon, M.J., and Castell, J.V. (2002). Down-regulation of human CYP3A4 by the inflammatory signal interleukin-6: molecular mechanism and transcription factors involved. *Faseb J* 16, 1799-1801.

Joyce, A.R., and Palsson, B.O. (2006). The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 7, 198-210.

Kacharina, J.E., Crino, P.B., and Eberwine, J. (1999). Preparation of cDNA from single cells and subcellular regions. *Methods Enzymol* 303, 3-18.

Kamiyama, Y., Matsubara, T., Yoshinari, K., Nagata, K., Kamimura, H., and Yamazoe, Y. (2007). Role of human hepatocyte nuclear factor 4alpha in the expression of drug-metabolizing enzymes and transporters in human hepatocytes assessed by use of small interfering RNA. *Drug Metab Pharmacokinet* 22, 287-298.

- Kane, M.D., Jatko, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., and Madore, S.J. (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28, 4552-4557.
- Kast, H.R., Goodwin, B., Tarr, P.T., Jones, S.A., Anisfeld, A.M., Stoltz, C.M., Tontonoz, P., Kliewer, S., Willson, T.M., and Edwards, P.A. (2002). Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J Biol Chem* 277, 2908-2915.
- Kitano, H. (2002). Standards for modeling. *Nat Biotechnol* 20, 337.
- Kliewer, S.A. (2003). The nuclear pregnane X receptor regulates xenobiotic detoxification. *J Nutr* 133, 2444S-2447S.
- Kliewer, S.A., and Willson, T.M. (2002). Regulation of xenobiotic and bile acid metabolism by the nuclear pregnane X receptor. *J Lipid Res* 43, 359-364.
- Knudsen, K.A., Frankowski, C., Johnson, K.R., and Wheelock, M.J. (1998). A role for cadherins in cellular signaling and differentiation. *J Cell Biochem Suppl* 30-31, 168-176.
- Kodama, S., Koike, C., Negishi, M., and Yamamoto, Y. (2004). Nuclear receptors CAR and PXR cross talk with FOXO1 to regulate genes that encode drug-metabolizing and gluconeogenic enzymes. *Mol Cell Biol* 24, 7931-7940.
- Kodama, S., Moore, R., Yamamoto, Y., and Negishi, M. (2007). Human nuclear pregnane X receptor cross-talk with CREB to repress cAMP activation of the glucose-6-phosphatase gene. *Biochem J* 407, 373-381.
- Koj, A. (1996). Initiation of acute phase response and synthesis of cytokines. *Biochim Biophys Acta* 1317, 84-94.
- Koop, D.R., Crump, B.L., Nordblom, G.D., and Coon, M.J. (1985). Immunochemical evidence for induction of the alcohol-oxidizing cytochrome P-450 of rabbit liver microsomes by diverse agents: ethanol, imidazole, trichloroethylene, acetone, pyrazole, and isoniazid. *Proc Natl Acad Sci U S A* 82, 4065-4069.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin,

J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrino, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Larkin, J.E., Frank, B.C., Gavras, H., Sultana, R., and Quackenbush, J. (2005). Independence and reproducibility across microarray platforms. *Nat Methods* 2, 337-344.

Le Phillip, P., Bahl, A., and Ungar, L.H. (2004). Using prior knowledge to improve genetic network reconstruction from microarray data. *In Silico Biol* 4, 335-353.

LeCluyse, E., Bullock, P., Madan, A., Carroll, K., and Parkinson, A. (1999). Influence of extracellular matrix overlay and medium formulation on the induction of cytochrome P-450 2B enzymes in primary cultures of rat hepatocytes. *Drug Metab Dispos* 27, 909-915.

LeCluyse, E., Madan, A., Hamilton, G., Carroll, K., DeHaan, R., and Parkinson, A. (2000). Expression and regulation of cytochrome P450 enzymes in primary cultures of human hepatocytes. *J Biochem Mol Toxicol* 14, 177-188.

- Lee, L.G., Connell, C.R., and Bloch, W. (1993). Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res* 21, 3761-3766.
- Lehmann, O.J., Sowden, J.C., Carlsson, P., Jordan, T., and Bhattacharya, S.S. (2003). Fox's in development and disease. *Trends Genet* 19, 339-344.
- Lemuth, K. (2006). Transkriptomanalyse von *Escherichia coli* unter Kohlenhydrat-Limitierung mittels DNA-Microarrays., University of Stuttgart, Stuttgart.
- Lettieri, T. (2006). Recent applications of DNA microarray technology to toxicology and ecotoxicology. *Environ Health Perspect* 114, 4-9.
- Li, D., Yang, X.L., Zhang, S.J., Lin, M., Yu, W.J., and Hu, K. (2007). Effects of mammalian CYP3A inducers on CYP3A-related enzyme activities in grass carp (*Ctenopharyngodon idellus*): Possible implications for the establishment of a fish CYP3A induction model. *Comp Biochem Physiol C Toxicol Pharmacol*.
- Li, S., Wu, L., and Zhang, Z. (2006). Constructing biological networks through combined literature mining and microarray analysis: a LMMA approach. *Bioinformatics* 22, 2143-2150.
- Li, T., and Chiang, J.Y. (2006). Rifampicin induction of CYP3A4 requires pregnane X receptor cross talk with hepatocyte nuclear factor 4alpha and coactivators, and suppression of small heterodimer partner gene expression. *Drug Metab Dispos* 34, 756-764.
- Lie, D.C., Colamarino, S.A., Song, H.J., Desire, L., Mira, H., Consiglio, A., Lein, E.S., Jessberger, S., Lansford, H., Dearie, A.R., and Gage, F.H. (2005). Wnt signalling regulates adult hippocampal neurogenesis. *Nature* 437, 1370-1375.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. *Nat Genet* 21, 20-24.
- Liu, J., Cheng, M.L., Shi, J.Z., Yang, Q., Wu, J., Li, C.X., and Waalkes, M.P. (2006). Differential effects between maotai and ethanol on hepatic gene expression in mice: Possible role of metallothionein and heme oxygenase-1 induction by maotai. *Exp Biol Med (Maywood)* 231, 1535-1541.
- Lönnstedt, I., and Speed, T. (2002). Replicated microarray data. *Statistica Sinica* 12 31 - 46.
- Luzzi, V.I., Holtschlag, V., and Watson, M.A. (2005). Gene expression profiling of primary tumor cell populations using laser capture microdissection, RNA transcript amplification, and GeneChip microarrays. *Methods Mol Biol* 293, 187-207.
- Lyamichev, V., Brow, M.A., and Dahlberg, J.E. (1993). Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260, 778-783.

- Mah, N., Thelin, A., Lu, T., Nikolaus, S., Kuhbacher, T., Gurbuz, Y., Eickhoff, H., Kloppel, G., Lehrach, H., Mellgard, B., Costello, C.M., and Schreiber, S. (2004). A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 16, 361-370.
- Makishima, M., Okamoto, A.Y., Repa, J.J., Tu, H., Learned, R.M., Luk, A., Hull, M.V., Lustig, K.D., Mangelsdorf, D.J., and Shan, B. (1999). Identification of a nuclear receptor for bile acids. *Science* 284, 1362-1365.
- Mandruzzato, S. (2007). Technological platforms for microarray gene expression profiling. *Adv Exp Med Biol* 593, 12-18.
- Marschall, H.U., Wagner, M., Zollner, G., Fickert, P., Diczfalusy, U., Gumhold, J., Silbert, D., Fuchsbichler, A., Benthin, L., Grundstrom, R., Gustafsson, U., Sahlin, S., Einarsson, C., and Trauner, M. (2005). Complementary stimulation of hepatobiliary transport and detoxification systems by rifampicin and ursodeoxycholic acid in humans. *Gastroenterology* 129, 476-485.
- Martignoni, M. (2006). Species and strain differences in drug metabolism in liver and intestine, University of Groningen, Groningen.
- Matsumoto, M., Poci, A., Rossetti, L., Depinho, R.A., and Accili, D. (2007). Impaired regulation of hepatic glucose production in mice lacking the forkhead transcription factor foxo1 in liver. *Cell Metab* 6, 208-216.
- Mauch, K., Müller, D., and Reuss, M. (2005). Integration von zellulären Stoffwechsel- und Signalnetzwerken. Universität Stuttgart: Stuttgart.
- McClain, K.L., Cai, Y.H., Hicks, J., Peterson, L.E., Yan, X.T., Che, S., and Ginsberg, S.D. (2005). Expression profiling using human tissues in combination with RNA amplification and microarray analysis: assessment of Langerhans cell histiocytosis. *Amino Acids* 28, 279-290.
- McGall, G.H., and Christians, F.C. (2002). High-density genechip oligonucleotide probe arrays. *Adv Biochem Eng Biotechnol* 77, 21-42.
- Mei, R., Hubbell, E., Bekiranov, S., Mittmann, M., Christians, F.C., Shen, M.M., Lu, G., Fang, J., Liu, W.M., Ryder, T., Kaplan, P., Kulp, D., and Webster, T.A. (2003). Probe selection for high-density oligonucleotide arrays. *Proc Natl Acad Sci U S A* 100, 11237-11242.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* 12, 7035-7056.
- Meunier, V., Bourrie, M., Julian, B., Marti, E., Guillou, F., Berger, Y., and Fabre, G. (2000). Expression and induction of CYP1A1/1A2, CYP2A6 and CYP3A4 in primary cultures of human hepatocytes: a 10-year follow-up. *Xenobiotica* 30, 589-607.

Moghe, P.V., Berthiaume, F., Ezzell, R.M., Toner, M., Tompkins, R.G., and Yarmush, M.L. (1996). Culture matrix configuration and composition in the maintenance of hepatocyte polarity and function. *Biomaterials* 17, 373-385.

Moncion, A., Truong, N.T., Garrone, A., Beaune, P., Barouki, R., and De Waziers, I. (2002). Identification of a 16-nucleotide sequence that mediates post-transcriptional regulation of rat CYP2E1 by insulin. *J Biol Chem* 277, 45904-45910.

Moore, L.B., Parks, D.J., Jones, S.A., Bledsoe, R.K., Consler, T.G., Stimmel, J.B., Goodwin, B., Liddle, C., Blanchard, S.G., Willson, T.M., Collins, J.L., and Kliewer, S.A. (2000). Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *J Biol Chem* 275, 15122-15127.

Nakae, J., Biggs, W.H., 3rd, Kitamura, T., Cavenee, W.K., Wright, C.V., Arden, K.C., and Accili, D. (2002). Regulation of insulin action and pancreatic beta-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1. *Nat Genet* 32, 245-253.

Nakae, J., Oki, M., and Cao, Y. (2007). The FoxO transcription factors and metabolic regulation. *FEBS Lett*.

Nakata, K., Tanaka, Y., Nakano, T., Adachi, T., Tanaka, H., Kaminuma, T., and Ishikawa, T. (2006). Nuclear receptor-mediated transcriptional regulation in Phase I, II, and III xenobiotic metabolizing systems. *Drug Metab Pharmacokinet* 21, 437-457.

Niemi, M., Backman, J.T., Fromm, M.F., Neuvonen, P.J., and Kivisto, K.T. (2003). Pharmacokinetic interactions with rifampicin : clinical relevance. *Clin Pharmacokinet* 42, 819-850.

Nizetic, D., Zehetner, G., Monaco, A.P., Gellen, L., Young, B.D., and Lehrach, H. (1991). Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc Natl Acad Sci U S A* 88, 3233-3237.

Pabon, C., Modrusan, Z., Ruvolo, M.V., Coleman, I.M., Daniel, S., Yue, H., and Arnold, L.J., Jr. (2001). Optimized T7 amplification system for microarray analysis. *Biotechniques* 31, 874-879.

Parks, D.J., Blanchard, S.G., Bledsoe, R.K., Chandra, G., Consler, T.G., Kliewer, S.A., Stimmel, J.B., Willson, T.M., Zavacki, A.M., Moore, D.D., and Lehmann, J.M. (1999). Bile acids: natural ligands for an orphan nuclear receptor. *Science* 284, 1365-1368.

Pascussi, J.M., Drocourt, L., Gerbal-Chaloin, S., Fabre, J.M., Maurel, P., and Vilarem, M.J. (2001). Dual effect of dexamethasone on CYP3A4 gene expression in human hepatocytes. Sequential role of glucocorticoid receptor and pregnane X receptor. *Eur J Biochem* 268, 6346-6358.

Pascussi, J.M., Gerbal-Chaloin, S., Drocourt, L., Assenat, E., Larrey, D., Pichard-Garcia, L., Vilarem, M.J., and Maurel, P. (2004). Cross-talk between xenobiotic

detoxication and other signalling pathways: clinical and toxicological consequences. *Xenobiotica* 34, 633-664.

Pauling, L., Itano, H.A., and et al. (1949). Sickle cell anemia a molecular disease. *Science* 110, 543-548.

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.

Podvinec, M., Handschin, C., Looser, R., and Meyer, U.A. (2004). Identification of the xenosensors regulating human 5-aminolevulinic synthase. *Proc Natl Acad Sci U S A* 101, 9127-9132.

Puigserver, P., Rhee, J., Donovan, J., Walkey, C.J., Yoon, J.C., Oriente, F., Kitamura, Y., Altomonte, J., Dong, H., Accili, D., and Spiegelman, B.M. (2003). Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1 α interaction. *Nature* 423, 550-555.

Quinn, P.G., and Yeagley, D. (2005). Insulin regulation of PEPCK gene expression: a model for rapid and reversible modulation. *Curr Drug Targets Immune Endocr Metabol Disord* 5, 423-437.

Reddy, A.B., Maywood, E.S., Karp, N.A., King, V.M., Inoue, Y., Gonzalez, F.J., Lilley, K.S., Kyriacou, C.P., and Hastings, M.H. (2007). Glucocorticoid signaling synchronizes the liver circadian transcriptome. *Hepatology* 45, 1478-1488.

Renton, K.W. (2005). Regulation of drug metabolism and disposition during inflammation and infection. *Expert Opin Drug Metab Toxicol* 1, 629-640.

Rodrigues, E., Vilarem, M.J., Ribeiro, V., Maurel, P., and Lechner, M.C. (2003). Two CCAAT/enhancer binding protein sites in the cytochrome P4503A1 locus. Potential role in the glucocorticoid response. *Eur J Biochem* 270, 556-564.

Ronaghi, M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome Res* 11, 3-11.

Roudkenar, M.H., Li, L., Baba, T., Kuwahara, Y., Nakagawa, H., Wang, L., Kasaoka, S., Ohkubo, Y., Ono, K., and Fukumoto, M. (2007). Gene Expression Profiles in Mouse Liver Cells after Exposure to Different Types of Radiation. *J Radiat Res (Tokyo)*.

Saghizadeh, M., Brown, D.J., Tajbakhsh, J., Chen, Z., Kenney, M.C., Farber, D.B., and Nelson, S.F. (2003). Evaluation of techniques using amplified nucleic acid probes for gene expression profiling. *Biomol Eng* 20, 97-106.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B., and Kinzler, K.W. (1997). Characterization of the yeast transcriptome. *Cell* 88, 243-251.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning - a laboratory manual*, 2nd ed. Cold Spring Harbor. Cold Spring Harbor Laboratory Press: New York.

Sarkar, J., Qi, C., Guo, D., Ahmed, M.R., Jia, Y., Usuda, N., Viswakarma, N., Rao, M.S., and Reddy, J.K. (2007). Transcription coactivator PRIP, the peroxisome proliferator-activated receptor (PPAR)-interacting protein, is redundant for the function of nuclear receptors PParalpha and CAR, the constitutive androstane receptor, in mouse liver. *Gene Expr* 13, 255-269.

Sauer, U., Heinemann, M., and Zamboni, N. (2007). Genetics. Getting closer to the whole picture. *Science* 316, 550-551.

Schadt, E.E., Li, C., Ellis, B., and Wong, W.H. (2001). Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J Cell Biochem Suppl Suppl* 37, 120-125.

Schena, M. (2003). *Microarray Analysis*. John Wiley & Sons, Inc. Hoboken, New Jersey, USA.

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.

Schindler, H., Wiese, A., Auer, J., and Burtscher, H. (2005). cRNA target preparation for microarrays: comparison of gene expression profiles generated with different amplification procedures. *Anal Biochem* 344, 92-101.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 7, 3.

Sekine, K., Chen, Y.R., Kojima, N., Ogata, K., Fukamizu, A., and Miyajima, A. (2007). Foxo1 links insulin signaling to C/EBPalpha and regulates gluconeogenesis during liver development. *Embo J* 26, 3607-3615.

Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., de Longueville, F., Kawasaki, E.S., Lee, K.Y., Luo, Y., Sun, Y.A., Willey, J.C., Setterquist, R.A., Fischer, G.M., Tong, W., Dragan, Y.P., Dix, D.J., Frueh, F.W., Goodsaid, F.M., Herman, D., Jensen, R.V., Johnson, C.D., Lobenhofer, E.K., Puri, R.K., Schrf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P.K., Zhang, L., Amur, S., Bao, W., Barbacioru, C.C., Lucas, A.B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X.M., Cebula, T.A., Chen, J.J., Cheng, J., Chu, T.M., Chudin, E., Corson, J., Corton, J.C., Croner, L.J., Davies, C., Davison, T.S., Delenstarr, G., Deng, X., Dorris, D., Eklund, A.C., Fan, X.H., Fang, H., Fulmer-Smentek, S., Fuscoe, J.C., Gallagher, K., Ge, W., Guo, L., Guo, X., Hager, J., Haje, P.K., Han, J., Han, T., Harbottle, H.C., Harris, S.C., Hatchwell, E., Hauser, C.A., Hester, S., Hong, H., Hurban, P., Jackson, S.A., Ji, H., Knight, C.R., Kuo, W.P., LeClerc, J.E., Levy, S., Li, Q.Z., Liu, C., Liu, Y., Lombardi, M.J., Ma, Y., Magnuson,

S.R., Maqsoodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B., Novoradovskaya, N., Orr, M.S., Osborn, T.W., Papallo, A., Patterson, T.A., Perkins, R.G., Peters, E.H., Peterson, R., Philips, K.L., Pine, P.S., Pusztai, L., Qian, F., Ren, H., Rosen, M., Rosenzweig, B.A., Samaha, R.R., Schena, M., Schroth, G.P., Shchegrova, S., Smith, D.D., Staedtler, F., Su, Z., Sun, H., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K.L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S.J., Wang, S.J., Wang, Y., Wolfinger, R., Wong, A., Wu, J., Xiao, C., Xie, Q., Xu, J., Yang, W., Zhang, L., Zhong, S., Zong, Y., and Slikker, W., Jr. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24, 1151-1161.

Shimamoto, Y., Ishida, J., Yamagata, K., Saito, T., Kato, H., Matsuoka, T., Hirota, K., Daitoku, H., Nangaku, M., Yamagata, K., Fujii, H., Takeda, J., and Fukamizu, A. (2004). Inhibitory effect of the small heterodimer partner on hepatocyte nuclear factor-4 mediates bile acid-induced repression of the human angiotensinogen gene. *J Biol Chem* 279, 7770-7776.

Sidhu, J.S., and Omiecinski, C.J. (1999). Insulin-mediated modulation of cytochrome P450 gene induction profiles in primary rat hepatocyte cultures. *J Biochem Mol Toxicol* 13, 1-9.

Silva, J.M., Morin, P.E., Day, S.H., Kennedy, B.P., Payette, P., Rushmore, T., Yergey, J.A., and Nicoll-Griffith, D.A. (1998). Refinement of an in vitro cell model for cytochrome P450 induction. *Drug Metab Dispos* 26, 490-496.

Sinal, C.J., Tohkin, M., Miyata, M., Ward, J.M., Lambert, G., and Gonzalez, F.J. (2000). Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* 102, 731-744.

Smyth, G.K., Yang, Y.H., and Speed, T. (2003). Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* 224, 111-136.

Song, B.J., Veech, R.L., Park, S.S., Gelboin, H.V., and Gonzalez, F.J. (1989). Induction of rat hepatic N-nitrosodimethylamine demethylase by acetone is due to protein stabilization. *J Biol Chem* 264, 3568-3572.

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.

Staudinger, J.L., Goodwin, B., Jones, S.A., Hawkins-Brown, D., MacKenzie, K.I., LaTour, A., Liu, Y., Klaassen, C.D., Brown, K.K., Reinhard, J., Willson, T.M., Koller, B.H., and Kliewer, S.A. (2001). The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc Natl Acad Sci U S A* 98, 3369-3374.

Stryer, L., and Haugland, R.P. (1967). Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A* 58, 719-726.

- Taga, T., and Kishimoto, T. (1997). Gp130 and the interleukin-6 family of cytokines. *Annu Rev Immunol* *15*, 797-819.
- Takeichi, M. (1988). The cadherins: cell-cell adhesion molecules controlling animal morphogenesis. *Development* *102*, 639-655.
- Takeichi, M. (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* *251*, 1451-1455.
- Tan, P.K., Downey, T.J., Spitznagel, E.L., Jr., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M., and Cam, M.C. (2003). Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* *31*, 5676-5684.
- Taylor, J.M., Illmensee, R., and Summers, J. (1976). Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. *Biochim Biophys Acta* *442*, 324-330.
- Tegude, H., Schnabel, A., Zanger, U.M., Klein, K., Eichelbaum, M., and Burk, O. (2007). Molecular mechanism of basal CYP3A4 regulation by hepatocyte nuclear factor 4alpha: evidence for direct regulation in the intestine. *Drug Metab Dispos* *35*, 946-954.
- Teng, S., and Piquette-Miller, M. (2005). The involvement of the pregnane X receptor in hepatic gene regulation during inflammation in mice. *J Pharmacol Exp Ther* *312*, 841-848.
- Terry, C., Hughes, R.D., Mitry, R.R., Lehec, S.C., and Dhawan, A. (2007). Cryopreservation-induced nonattachment of human hepatocytes: role of adhesion molecules. *Cell Transplant* *16*, 639-647.
- Tirona, R.G., Lee, W., Leake, B.F., Lan, L.B., Cline, C.B., Lamba, V., Parviz, F., Duncan, S.A., Inoue, Y., Gonzalez, F.J., Schuetz, E.G., and Kim, R.B. (2003). The orphan nuclear receptor HNF4alpha determines PXR- and CAR-mediated xenobiotic induction of CYP3A4. *Nat Med* *9*, 220-224.
- Tompkins, L.M., and Wallace, A.D. (2007). Mechanisms of cytochrome P450 induction. *J Biochem Mol Toxicol* *21*, 176-181.
- Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* *98*, 5116-5121.
- Upton, J.J., Stoyanova, R., Cooper, H.S., Patriotis, C., Ross, E.A., Boman, B., Clapper, M.L., Knudson, A.G., and Bellacosa, A. (2004). Optimized procedures for microarray analysis of histological specimens processed by laser capture microdissection. *J Cell Physiol* *201*, 366-373.
- 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.

van Someren, E.P., Wessels, L.F., Backer, E., and Reinders, M.J. (2002). Genetic network modeling. *Pharmacogenomics* 3, 507-525.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270, 484-487.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. *Science* 291, 1304-1351.

von Bertalanffy, L. (1951). General system theory; a new approach to unity of science. 1. Problems of general system theory. *Hum Biol* 23, 302-312.

Wang, H., Chen, J., Hollister, K., Sowers, L.C., and Forman, B.M. (1999). Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Mol Cell* 3, 543-553.

Wang, H., Faucette, S.R., Gilbert, D., Jolley, S.L., Sueyoshi, T., Negishi, M., and LeCluyse, E.L. (2003). Glucocorticoid receptor enhancement of pregnane X receptor-mediated CYP2B6 regulation in primary human hepatocytes. *Drug Metab Dispos* 31, 620-630.

Wang, H., and LeCluyse, E.L. (2003). Role of orphan nuclear receptors in the regulation of drug-metabolising enzymes. *Clin Pharmacokinet* 42, 1331-1357.

Wang, J.C., Stafford, J.M., and Granner, D.K. (1998). SRC-1 and GRIP1 coactivate transcription with hepatocyte nuclear factor 4. *J Biol Chem* 273, 30847-30850.

Wang, S., and Cheng, Q. (2006). Microarray analysis in drug discovery and clinical applications. *Methods Mol Biol* 316, 49-65.

Wang, Z., Qi, C., Kronos, A., Woodring, P., Zhu, X., Reddy, J.K., Evans, R.M., Rosenfeld, M.G., and Hunter, T. (2006). Critical roles of the p160 transcriptional coactivators p/CIP and SRC-1 in energy balance. *Cell Metab* 3, 111-122.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.

Waxman, D.J. (1999). P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. *Arch Biochem Biophys* 369, 11-23.

Wiegant, J., Verwoerd, N., Mascheretti, S., Bolk, M., Tanke, H.J., and Raap, A.K. (1996). An evaluation of a new series of fluorescent dUTPs for fluorescence in situ hybridization. *J Histochem Cytochem* 44, 525-529.

Wojcikowski, J., Golembiowska, K., and Daniel, W.A. (2007). The regulation of liver cytochrome p450 by the brain dopaminergic system. *Curr Drug Metab* 8, 631-638.

Xiang, C.C., Chen, M., Kozhich, O.A., Phan, Q.N., Inman, J.M., Chen, Y., and Brownstein, M.J. (2003). Probe generation directly from small numbers of cells for DNA microarray studies. *Biotechniques* 34, 386-388, 390, 392-383.

Yang, Y.H., and Speed, T. (2002). Design issues for cDNA microarray experiments. *Nat Rev Genet* 3, 579-588.

Yoon, J.C., Puigserver, P., Chen, G., Donovan, J., Wu, Z., Rhee, J., Adelman, G., Stafford, J., Kahn, C.R., Granner, D.K., Newgard, C.B., and Spiegelman, B.M. (2001). Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* 413, 131-138.

Yoshitsugu, H., Nishimura, M., Tateno, C., Kataoka, M., Takahashi, E., Soeno, Y., Yoshizato, K., Yokoi, T., and Naito, S. (2006). Evaluation of human CYP1A2 and CYP3A4 mRNA expression in hepatocytes from chimeric mice with humanized liver. *Drug Metab Pharmacokinet* 21, 465-474.

Zhang, W., Patil, S., Chauhan, B., Guo, S., Powell, D.R., Le, J., Klotsas, A., Matika, R., Xiao, X., Franks, R., Heidenreich, K.A., Sajan, M.P., Farese, R.V., Stolz, D.B., Tso, P., Koo, S.H., Montminy, M., and Unterman, T.G. (2006). FoxO1 regulates multiple metabolic pathways in the liver: effects on gluconeogenic, glycolytic, and lipogenic gene expression. *J Biol Chem* 281, 10105-10117.

Zilly, W., Breimer, D.D., and Richter, E. (1977). Pharmacokinetic interactions with rifampicin. *Clin Pharmacokinet* 2, 61-70.

Zinn, K., DiMaio, D., and Maniatis, T. (1983). Identification of two distinct regulatory regions adjacent to the human beta-interferon gene. *Cell* 34, 865-879.

Zollner, G., and Trauner, M. (2006). Molecular mechanisms of cholestasis. *Wien Med Wochenschr* 156, 380-385.

<http://www.montp.inserm.fr/u632/>

<http://www.ncbi.nlm.nih.gov/>

6 Appendix

Table 58: At least 2-fold up-regulated genes in primary hepatocytes after 6 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 6 h of induction.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	CYP26A1	cytochrome P450, family 26, subfamily A, polypeptide 1	6.5	1.0	1.0
2	MGC39372	hypothetical protein MGC39372	5.4	9.5	3.8
3	LOC343071	similar to Hypothetical protein DJ845O24.1	3.9	3.1	5.4
4	THRSP	thyroid hormone responsive (SPOT14 homolog, rat)	3.7	1.7	1.0
5	TNRC9	trinucleotide repeat containing 9	3.1	3.0	4.1
6	FASN	fatty acid synthase	2.8	1.4	0.5
7	SH3TC1	SH3 domain and tetratricopeptide repeats 1	2.8	1.7	1.1
8	DLG7	discs large homolog 7	2.2	1.4	1.2
9	ARHGDI A	Rho GDP dissociation inhibitor (GDI) alpha	2.2	0.9	0.6
10	ALAS1	5-aminolevulinatase synthase	2.2	2.9	2.6
11	LOC440995	hypothetical gene supported by BC034933; BC068085	2.1	1.3	1.1
12	BCLP	beta-casein-like protein	2.1	1.6	1.2
13	OKL38	pregnancy-induced growth inhibitor	2.0	1.4	0.8
14	FLJ13111	hypothetical protein FLJ13111	2.0	1.2	0.6
15	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	2.0	3.7	3.0
16	CTSZ	CTSZ protein	2.0	0.5	1.0

Table 59: At least 2-fold down-regulated genes in primary hepatocytes after 6 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 6 h of induction.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	ADH1B	alcohol dehydrogenase (class I), beta polypeptide	0.3	0.1	0.1
2	SLC38A4	solute carrier family 38, member 4	0.4	0.5	0.5
3	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	0.4	0.3	0.8
4	CXCL10	CXC chemokine ligand 10	0.4	0.6	1.2
5	LOC283680	hypothetical protein LOC283680	0.5	0.6	0.8
6	GLS2	glutaminase 2 (liver, mitochondrial)	0.5	0.3	0.5
7	LOC221362	hypothetical protein LOC221362	0.5	1.3	1.2
8	MGC9084	hypothetical protein MGC9084	0.5	0.8	1.1
9	SLC6A1	solute carrier family 6 (neurotransmitter transporter, GABA), member 1	0.5	0.4	0.9
10	GPR64	G protein-coupled receptor 64	0.5	1.0	0.1
11	HTR2B	5-hydroxytryptamine (serotonin) receptor 2B	0.5	0.3	0.7
12	ADH4	Adh4p	0.5	0.4	0.5
13	LOC440858	LOC440858	0.5	0.7	0.8
14	ADH6	Adh6p	0.5	0.4	0.5

Table 60: At least 2-fold up-regulated genes in primary hepatocytes after 24 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 24 h of induction.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	MGC39372	hypothetical protein MGC39372	5.4	9.5	3.8
2	STK39	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)	1.5	3.8	2.4
3	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	2.0	3.7	3.0
4	MT1F	metallothionein 1F (functional)	1.2	3.2	0.9
5	LOC343071	similar to Hypothetical protein DJ845O24.1	3.9	3.1	5.4
6	TNRC9	trinucleotide repeat containing 9	3.1	3.0	4.1

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
7	ALAS1	5-aminolevulinatase synthase	2.2	2.9	2.6
8	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	1.6	2.9	2.9
9	PRODH2	proline dehydrogenase (oxidase) 2	1.6	2.8	2.0
10	SEC14L4	SEC14-like 4 (<i>S. cerevisiae</i>)	1.9	2.8	2.0
11	UPP1	uridine phosphorylase 1	1.9	2.7	1.3
12	MT1X	metallothionein 1X	1.0	2.6	0.8
13	CCL2	chemokine (C-C motif) ligand 2	1.3	2.6	3.2
14	AGXT2L1	alanine-glyoxylate aminotransferase 2-like 1	1.6	2.6	1.8
15	NPM3	nucleophosmin/nucleoplasmin, 3	1.2	2.4	1.2
16	FGF19	fibroblast growth factor 19	1.6	2.4	1.9
17	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	1.0	2.3	2.1
18	GDF15	growth differentiation factor 15	1.6	2.3	2.3
19	MT1K	metallothionein 1K	1.3	2.3	0.8
20	CYP3A43	cytochrome P450, family 3, subfamily A, polypeptide 43	1.2	2.3	2.1
21	CYP2A6	cytochrome P450, family 2, subfamily A, polypeptide 6	0.9	2.3	1.0
22	ABCB1	ATP-binding cassette sub-family B member 1	1.3	2.2	2.4
23	KIAA1702	KIAA1702 protein	1.0	2.2	1.5
24	CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8	1.7	2.2	4.7
25	BTBD11	BTB (POZ) domain containing 11	1.4	2.2	1.6
26	CHC1	Chc1p	1.0	2.1	1.5
27	SGK2	serum/glucocorticoid regulated kinase 2	1.6	2.1	1.3
28	CYP2A7	cytochrome P450, family 2, subfamily A, polypeptide 7	0.8	2.1	0.9
29	MSX1	homeobox containing peptide Xhox 7.1	1.6	2.1	1.4
30	CA12	carbonic anhydrase XII	1.2	2.1	2.0
31	CD14	CD14 antigen	1.2	2.0	1.9
32	WRN	Werner syndrome	0.8	2.0	2.0
33	TGFB2	transforming growth factor beta 2	1.1	2.0	1.8
34	MAP2	Map2p	1.1	2.0	0.6

Table 61: At least 2-fold down-regulated genes in primary hepatocytes after 24 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 24 h of induction.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	CYP7A1	cytochrome P450 7A1	0.6	0.1	0.2
2	ADH1A	alcohol dehydrogenase (class I), alpha polypeptide	0.7	0.1	0.2
3	ADH1B	alcohol dehydrogenase (class I), beta polypeptide	0.3	0.1	0.1
4	ADH1C	alcohol dehydrogenase 1C (class 1), gamma polypeptide	0.6	0.2	0.2
5	HTR2B	5-hydroxytryptamine (serotonin) receptor 2B	0.5	0.3	0.7
6	GLS2	glutaminase 2 (liver, mitochondrial)	0.5	0.3	0.5
7	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	0.4	0.3	0.8
8	CYP2E1	cytochrome P450 2E1	0.7	0.3	0.2
9	ASPA	ASPARTATE AMMONIA-LYASE	0.6	0.3	1.0
10	C6orf142	chromosome 6 open reading frame 142	0.7	0.3	0.5
11	MGC45562	hypothetical protein MGC45562	0.6	0.4	0.7
12	PGLYRP2	peptidoglycan recognition protein 2	0.7	0.4	0.4
13	AMY2A	amylase, alpha 2A; pancreatic	0.7	0.4	0.5
14	PKLR	pyruvate kinase, liver and RBC	1.0	0.4	0.4
15	CYP1A2	cytochrome P450 1A2	0.7	0.4	0.6
16	COL5A2	alpha 2 type V collagen	0.6	0.4	0.7
17	ADH4	Adh4p	0.5	0.4	0.5
18	LOC151438	hypothetical protein LOC151438	0.8	0.4	0.6
19	SPARCL1	SPARC-like 1 (mast9, hevin)	0.9	0.4	0.7
20	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	0.7	0.4	0.6
21	SPP2	promotes the first step of splicing and is required for spliceosome maturation	0.6	0.4	0.5
22	RODH-4		0.7	0.4	0.2
23	GPR88	G-protein coupled receptor 88	0.6	0.4	0.4
24	SLC6A1	solute carrier family 6 (neurotransmitter transporter, GABA), member 1	0.5	0.4	0.9

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
25	CPN1	carboxypeptidase N, polypeptide 1, 50kD	0.7	0.4	0.7
26	AMIGO2	amphoterin induced gene 2	0.7	0.4	1.2
27	LOC255167	hypothetical protein LOC255167	0.7	0.4	0.6
28	HCP5	HLA complex P5	0.5	0.4	1.2
29	PCK1	Pck1p	0.9	0.4	1.4
30	GUCA2B	guanylate cyclase activator 2B (uroguanylin)	1.1	0.4	0.7
31	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	0.6	0.4	0.8
32	KCNN2	potassium calcium-activated channel, subfamily N, member 2	0.7	0.4	0.9
33	KIAA1305	KIAA1305	1.0	0.4	0.9
34	FAM44A	family with sequence similarity 44, member A	0.6	0.4	1.2
35	ZGPAT	zinc finger, CCCH-type with G patch domain	1.0	0.4	0.5
36	AMT	aminomethyltransferase (glycine cleavage system protein T)	0.7	0.4	0.6
37	MRC1	Mrc1p	0.5	0.4	0.5
38	HAO2	hydroxyacid oxidase 2 (long chain)	0.6	0.4	0.1
39	OLFML1	olfactomedin-like 1	0.9	0.4	0.8
40	FLJ20581	hypothetical protein FLJ20581	0.7	0.4	0.7
41	ZMYND12	zinc finger, MYND-type containing 12	0.6	0.4	0.8
42	IGF1	insulin like growth factor 1	0.7	0.4	0.7
43	NR1I3	nuclear receptor subfamily 1, group I, member 3	0.6	0.4	0.5
44	ADH6	Adh6p	0.5	0.4	0.5
45	LOX	lysyl oxidase	1.6	0.4	0.9
46	SERPINA10	serine proteinase inhibitor, clade A member 10	0.7	0.5	0.7
47	AFM	afamin	0.7	0.5	0.4
48	H19	H19 fetal liver mRNA	1.7	0.5	1.0
49	GPR54	G protein-coupled receptor 54	1.5	0.5	1.4
50	UNQ830	ASCL830	0.8	0.5	0.6
51	APG4C	APG4 autophagy 4 homolog C (<i>S. cerevisiae</i>)	0.8	0.5	1.0
52	SNAI2	snail homolog 2 (<i>Drosophila</i>)	0.6	0.5	0.4

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
53	IFIT2	interferon-induced protein with tetratricopeptide repeats 2	0.8	0.5	0.9
54	NUCB1	CG32190-PA	1.3	0.5	0.8
55	FMO1	Flavin-containing monooxygenase, ER membrane; catalyzes oxidation	1.1	0.5	1.1
56	OGDHL	oxoglutarate dehydrogenase-like	0.6	0.5	0.7
57	SAH	SA hypertension-associated homolog (rat)	0.6	0.5	0.9
58	C6orf114	chromosome 6 open reading frame 114	0.5	0.5	1.0
59	FIS	FIS	0.6	0.5	0.3
60	LOC492304	putative insulin-like growth factor II associated protein	1.2	0.5	0.5
61	C14orf94	chromosome 14 open reading frame 94	0.8	0.5	0.6
62	LOC149703	hypothetical protein LOC149703	0.7	0.5	0.3
63	GYS2	glycogen synthase 2 (liver)	0.7	0.5	0.2
64	KLKB1	kallikrein	0.8	0.5	0.7

Table 62: At least 2-fold up-regulated genes in primary hepatocytes after 72 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 72 h of induction.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	LOC343071	similar to Hypothetical protein DJ845O24.1	3.9	3.1	5.4
2	CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8	1.7	2.2	4.7
3	TNRC9	trinucleotide repeat containing 9	3.1	3.0	4.1
4	NPPB	natriuretic peptide precursor B	1.5	1.6	4.0
5	MGC39372	hypothetical protein MGC39372	5.4	9.5	3.8
6	CCL2	chemokine (C-C motif) ligand 2	1.3	2.6	3.2
7	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	2.0	3.7	3.0
8	TLR1	Toll-like receptor 1	1.5	1.4	3.0
9	LMCD1	LIM and cysteine-rich domains 1	0.8	1.4	2.9
10	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	1.6	2.9	2.9

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
11	CRP	C-reactive protein	1.2	1.4	2.7
12	SLC13A5	solute carrier family 13 (citrate transporter), member 5	1.2	1.5	2.6
13	ALAS1	5-aminolevulinate synthase	2.2	2.9	2.6
14	FLJ22595		0.8	1.2	2.6
15	AKR1D1	aldo-keto reductase family 1, member D1	1.2	1.8	2.6
16	LOC93082	hypothetical protein BC012317	1.6	1.8	2.5
17	NR1D2	nuclear receptor subfamily 1, group D, member 2	0.8	0.8	2.5
18	FST	follicle-stimulating hormone receptor	1.3	1.1	2.4
19	ABCB1	ATP-binding cassette sub-family B member 1	1.3	2.2	2.4
20	STK39	serine/threonine kinase 39 (STE20/SPS1 homolog, yeast)	1.5	3.8	2.4
21	RRAD	Ras-related associated with diabetes	1.1	1.2	2.3
22	CA3	carbonic anhydrase III	1.2	1.0	2.3
23	GDF15	growth differentiation factor 15	1.6	2.3	2.3
24	PMAIP1	TNFRSF25	1.0	1.0	2.3
25	NMES1	normal mucosa of esophagus specific 1	1.5	1.1	2.3
26	MGC10850	hypothetical protein MGC10850	1.2	1.7	2.2
27	CYP3A7	cytochrome P450, family 3, subfamily A, polypeptide 7	1.2	1.8	2.1
28	SOX9	SOX9 (sex determining region Y)-box 9	1.6	1.9	2.1
29	BDKRB2	B2 bradykinin receptor	1.3	1.2	2.1
30	IGFBP3	insulin-like growth factor binding protein 3	1.4	0.9	2.1
31	CYP3A43	cytochrome P450, family 3, subfamily A, polypeptide 43	1.2	2.3	2.1
32	LBH	likely ortholog of mouse limb-bud and heart gene	1.4	1.2	2.1
33	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	1.0	2.3	2.1
34	C9orf19	chromosome 9 open reading frame 19	1.1	1.1	2.0
35	SNARK		1.0	1.9	2.0
36	PRODH2	proline dehydrogenase (oxidase) 2	1.6	2.8	2.0
37	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	1.2	1.4	2.0
38	MGAM	maltase-glucoamylase	1.0	1.0	2.0

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
39	CA12	carbonic anhydrase XII	1.2	2.1	2.0
40	SERPINE2	plasminogen activator inhibitor type 1, member 2	1.0	1.2	2.0

Table 63: At least 2-fold down-regulated genes in primary hepatocytes after 72 h of Rifampicin induction. Given are the gene name and description as well as the n-fold regulation score for the three investigated time points (6, 24 and 72 h). The data is ordered by the rate of fold-change after 72 h of induction.

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
1	LOC400793	hypothetical gene supported by AK125122	1.0	0.9	0.1
2	ADH1B	alcohol dehydrogenase (class I), beta polypeptide	0.3	0.1	0.1
3	GPR64	G protein-coupled receptor 64	0.5	1.0	0.1
4	HAO2	hydroxyacid oxidase 2 (long chain)	0.6	0.4	0.1
5	CYP7A1	cytochrome P450 7A1	0.6	0.1	0.2
6	ADH1A	alcohol dehydrogenase (class I), alpha polypeptide	0.7	0.1	0.2
7	GYS2	glycogen synthase 2 (liver)	0.7	0.5	0.2
8	CYP2E1	cytochrome P450 2E1	0.7	0.3	0.2
9	ADH1C	alcohol dehydrogenase 1C (class 1), gamma polypeptide	0.6	0.2	0.2
10	RODH-4		0.7	0.4	0.2
11	FIS	FIS	0.6	0.5	0.3
12	BHMT	betaine-homocysteine methyltransferase	0.8	0.6	0.3
13	LOC149703	hypothetical protein LOC149703	0.7	0.5	0.3
14	LIPG	endothelial lipase	0.8	0.6	0.3
15	GLYAT	glycine-N-acyltransferase	0.7	0.6	0.3
16	FKBP5	FK506 binding protein 5	0.7	0.8	0.3
17	PEG10	paternally expressed 10	1.0	0.9	0.3
18	CYP8B1	cytochrome P-450 8B1	0.9	0.5	0.3
19	GNMT	glycine N-methyltransferase	0.6	0.5	0.4
20	NPR3	natriuretic peptide receptor C	0.8	1.0	0.4
21	SLC22A7	solute carrier family 22 (organic anion transporter), member 7	0.9	0.5	0.4

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
22	SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	0.9	0.6	0.4
23	PCYT2	phosphate cytidyltransferase 2, ethanolamine	1.4	1.1	0.4
24	HAMP	Ham percentage	0.7	0.7	0.4
25	TGFBR3	transforming growth factor, beta receptor III (betaglycan, 300kDa)	0.8	0.9	0.4
26	C14orf58	chromosome 14 open reading frame 58	1.1	1.0	0.4
27	SNAI2	snail homolog 2 (Drosophila)	0.6	0.5	0.4
28	SLC38A3	solute carrier family 38, member 3	1.1	0.6	0.4
29	APOC4	apolipoprotein C-IV	0.8	1.0	0.4
30	C20orf121	chromosome 20 open reading frame 121	0.7	0.8	0.4
31	EPHX2	epoxide hydrolase 2, cytoplasmic	0.8	0.5	0.4
32	FLJ14054	hypothetical protein FLJ14054	0.9	1.2	0.4
33	PKLR	pyruvate kinase, liver and RBC	1.0	0.4	0.4
34	SLC22A1	solute carrier family 22 (organic cation transporter), member 1	1.0	0.7	0.4
35	AFM	afamin	0.7	0.5	0.4
36	LOC401081		0.7	0.7	0.4
37	ETNK2	ethanolamine kinase 2	0.9	0.7	0.4
38	FLJ20406		1.3	0.6	0.4
39	PER1	Per1p	1.1	1.0	0.4
40	GPR88	G-protein coupled receptor 88	0.6	0.4	0.4
41	PGLYRP2	peptidoglycan recognition protein 2	0.7	0.4	0.4
42	MGC3101	hypothetical protein MGC3101	1.1	0.6	0.5
43	KIAA1505	KIAA1505 protein	0.8	0.8	0.5
44	ANGPTL7	angiopoietin-like 7	1.0	0.7	0.5
45	RORC	RAR-related orphan receptor C	0.9	0.8	0.5
46	GPT	glutamic-pyruvate transaminase	1.0	0.8	0.5
47	LOC286044	hypothetical protein LOC286044	0.7	0.6	0.5
48	MOGAT3	monoacylglycerol O-acyltransferase 3	1.4	0.7	0.5
49	LDHD	lactate dehydrogenase D	1.0	0.6	0.5

No.	Gene Name	Description	6h N-Fold	24h N-Fold	72h N-Fold
50	EPA1	EPH receptor A1	0.7	0.8	0.5
51	C6	clone C6	0.8	0.7	0.5
52	LOC492304	putative insulin-like growth factor II associated protein	1.2	0.5	0.5
53	MRC1	Mrc1p	0.5	0.4	0.5
54	MGC32871	hypothetical protein MGC32871	0.6	0.6	0.5
55	GBA3	glucosidase, beta, acid 3 (cytosolic)	0.6	0.7	0.5
56	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	0.7	0.7	0.5
57	ZGPAT	zinc finger, CCCH-type with G patch domain	1.0	0.4	0.5
58	SLC10A1	solute carrier family 10 (sodium/bile acid cotransporter family), member 1	0.9	0.7	0.5
59	IGSF8	immunoglobulin superfamily, member 8	1.7	0.7	0.5
60	SLC38A4	solute carrier family 38, member 4	0.4	0.5	0.5
61	HAAO	3-hydroxyanthranilate 3,4-dioxygenase	0.9	0.6	0.5
62	CHST2	carbohydrate (N-acetylglucosamine-6-O) sulfotransferase 2	0.9	0.9	0.5
63	HSD3B1	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	0.9	0.7	0.5
64	DHFR	dihydrofolate reductase	1.0	0.7	0.5
65	MGC4399	mitochondrial carrier protein	1.0	0.9	0.5
66	MOGAT2	monoacylglycerol O-acyltransferase 2	0.8	0.6	0.5
67	LOC442096	similar to Kruppel-like factor 7 (ubiquitous); transcription factor	0.7	0.9	0.5
68	MME	membrane metallo-endopeptidase	0.8	0.7	0.5
69	LOC197350	hypothetical protein LOC197350	1.1	0.8	0.5
70	FBXO2	F-box protein 2	1.0	0.9	0.5

Genes investigated in the sub-genome analysis are given in Table 64:

Table 64: Gene table showing genes used for sub-genome expression analysis using Eppendorf hepato DualChips. Given are the number, gene name, gene symbol, Genbank and Swissprot accession numbers and the general biological function of the gene.

No.	Gene Name	Official Gene Symbol	Genbank	Swissprot	General Function
1	AcylCoenzyme A oxidase	ACOX1	S69189	Q15067	Lipid metabolism
2	Ahr	AHR	NM_001621	P35869	Transcription
3	Alpha-2-macroglobulin	A2M	NM_000014	P01023	circulation
4	Apo E	APOE	M12529	P02649	Lipid metabolism
5	ApolipoproteinJ	CLU	J02908	P10909	Lipid metabolism
6	BCRP	ABCG2	NM_004827	Q9UNQ0	Transport
7	BCL2-antagonist of cell death	BAD	NM_004322	Q92934	Apoptosis
8	BCL2-associated X protein	BAX	NM_004324	Q8WXU1	Apoptosis
9	B-cell lymphoma2	BCL2	NM_000633	P10415	Apoptosis
10	BCLX	BCLX	NM_001191	Q07817	Apoptosis
11	23KDa Highly basic protein	RPL13A	X56932	P39023	HouseKeeping gene
12	ATP-binding cassette, sub-family B (MDR/TAP)	ABCB11	NM_003742	Q95342	Multidrug Resistance Transporter
13	c-fos	FOS	NM_005252	P01100	Oncogenesis
14	c-jun	MAP2K7	AF013589	O14733	Stress response
15	Aldolase A, fructose biphosphate	ALDOA	NM_000034	P04075	HouseKeeping gene
16	c-myc	MYCBP	NM_012333	Q99417	Oncogenesis
17	Cadherin 12	CDH12	NM_004061	P55289	Cell adhesion
18	Constitutive androstane receptor	NR1I3	NM_005122	Q14994	Transcription
19	Caspase7	CASP7	NM_001227	P55210	Apoptosis
20	Caspase8	CASP8	X98172	O15519	Apoptosis
21	CCAAT/enhancer-binding protein	CEBPA	NM_004364	P49715	Transcription
22	Tumor necrosis factor receptor	TNFRSF7	NM_001242	P26842	Cell signaling / receptor
23	collagenVI-alpha2	COL6A2	NM_001849	P12110	Extracellular matrix
24	Catechol-O-methyltransferase	COMT	NM_000754	P21964	Protein synthesis
25	Prostaglandin endoperoxidase synthase 2	PTGS2	NM_000963	P35354	Lipid metabolism
26	cyclinA	CCNA1	NM_003914	P78396	Cell cycle
27	cyclinB1	CCNB1	NM_031966	P14635	Cell cycle
28	cyclinD1	CCND1	NM_053056	P24385	Cell cycle
29	cyclinD2	CCND2	NM_001759	P30279	Cell cycle
30	Alpha-tubulin	K-ALPHA-1	NM_006082	P05209	HouseKeeping gene
31	cyclinD3	CCND3	NM_001760	P30281	Cell cycle
32	cyclinE1	CCNE1	NM_001238	P24864	Cell cycle
33	cyclin-dependent kinase 2	cdk 2	NM_001798	P24941	Cell cycle
34	cyclin-dependent kinase 4	cdk 4	U79269	P11802	Cell cycle
35	cyclin-dependent kinase 6	cdk 6	NM_001259	Q00534	Cell cycle
36	Beta-Actin	ACTB	NM_001101	P02570	HouseKeeping gene
37	Cytochrome P450 1A1	CYP1A1	NM_000499	P04798	Drug metabolism
38	Cytochrome P450 1A2	CYP1A2	NM_000761	P05177	Drug metabolism
39	Cytochrome P450 1B1	CYP1B1	NM_000104	Q16678	Drug metabolism
40	Cytochrome P450 2A6/7/13	CYP2A	NM_000762	P11509	Drug metabolism
41	Cytochrome P450 2B6	CYP2B6	NM_000767	P20813	Drug metabolism
42	Cytochrome P450 2C8/9/19	CYP2C	M61854	P33261	Drug metabolism
43	Cytochrome P450 2D6	CYP2D6	X16866	P10635	Drug metabolism

Table 64 continued:

No.	Gene Name	Official Gene Symbol	Genbank	Swissprot	General Function
44	Cytochrome P450 2E1	CYP2E1	NM_000773	P05181	Drug metabolism
45	Cytochrome P450 3A3/4/5/7	CYP3A	NM_017460	P08684	Drug metabolism
46	Cytochrome P450 4A11	CYP4A11	NM_000778	Q02928	Drug metabolism
47	Cytochrome P450 4B1	CYP4B1	NM_000779	P13584	Drug metabolism
48	Cytochrome P450 4F3	CYP4F3	NM_000896	Q08477	Drug metabolism
49	Cytochrome P450 2C18	CYP2C18	M61853	P33260	Drug metabolism
50	E-cadherine	CDH1	NM_004360	P12830	Cell adhesion
51	elk-1	ELK1	NM_005229	P19419	Oncogenesis
52	enoyl CoA	EHHADH	NM_001966	Q08426	Lipid metabolism
53	Epoxide hydrolase microsomal	EPHX1	NM_000120	P07099	Drug metabolism
54	Epoxide hydrolase soluble	EPHX2	NM_001979	P34913	Drug metabolism
55	Cyclophilin 33 A	PPIE	AF042385	Q9UNP9	HouseKeeping gene
56	Mitogen activated protein kinase kinase 1	MAP2K1	NM_002755	Q02750	Cell cycle
57	fibronectin	FN1	M10905	P02751	Extracellular matrix
58	flavin containing monooxygenase 4	FMO4	NM_002022	P31512	Drug metabolism
59	flavin containing monooxygenase 5	FMO5	NM_001461	P49326	Drug metabolism
60	GADD153	DDIT3	S40706	P35638	DNA repair / synthesis
61	GADD45	GADD45A	NM_001924	P24522	DNA repair / synthesis
62	Glutathione reductase	GSR	NM_000637	P00390	Drug metabolism
63	Glutathione synthetase	GSS	NM_000178	P48637	Drug metabolism
64	glutathione peroxidase	GPX1	M21304	P07203	Oxidative metabolism
65	GST A1/2/3/4	GSTA	NM_000846	P09210	Drug metabolism
66	GST K1	LOC51064	NM_015917	Q9Y2Q3	Drug metabolism
67	GST M1/2	GSTM1	NM_000561	P09488	Drug metabolism
68	GADPH	GAPD	NM_002046	P04406	HouseKeeping gene
69	GST M3	GSTM3	NM_000849	P21266	Drug metabolism
70	GST M5	GSTM5	NM_000851	P46439	Drug metabolism
71	GST O1	GSTTLp28	NM_004832	P78417	Drug metabolism
72	GST T1	GSTT1	NM_000853	P30711	Drug metabolism
73	GST T2	GSTT2	NM_000854	P30712	Drug metabolism
74	GST Z1	GSTZ1	NM_001513	O43708	Drug metabolism
75	Glutathione S-transferase pi	GSTP1	NM_000852	P09211	Drug metabolism
76	GRLF1	GRLF1	NM_004491	Q9NRY4	Transcription
77	heme oxygenase	HMOX1	NM_002133	P09601	Stress response
78	hexokinase	HK1	M75126	P19367	HouseKeeping gene
79	hepatocyte growth factor (hepapoietin A)	HGF	X16323	P14210	Growth factor and cytokines
80	hepatocyte growth factor receptor	MET	NM_000245	P08581	Cell signaling / receptor
81	Hypoxanthine phosphoribosyltransferase	HPRT1	NM_000194	P00492	HouseKeeping gene
82	hepatocyte nuclear factor 4, alpha	HNF4A	NM_000457	P41235	Lipid metabolism
83	HGF activator	HGFAC	NM_001528	Q04756	Growth factor and cytokines
84	Heat shock 27kD protein1	HSPB1	AB020027	P04792	Stress response
85	Heat shock 40kD protein1	DNAJB1	D49547	P25685	Stress response
86	Heat shock 70kD protein1	HSPA4	AB023420	P34932	Stress response
87	Insulin like growth factor1	IGF1	X57025	P01343	Growth factor and cytokines
88	Insulin like growth factor1 receptor	IGF1R	NM_000875	P08069	Cell signaling / receptor
89	Insulin growth factor binding protein2	IGFBP2	M35410	P18065	Growth factor and cytokines
90	Insulin growth factor binding protein3	IGFBP3	X64875	P17936	Growth factor and cytokines
91	Insulin growth factor binding protein5	IGFBP5	M65062	P24593	Growth factor and cytokines

Table 64 continued:

No.	Gene Name	Official Gene Symbol	Genbank	Swissprot	General Function
92	IKB-alpha	NFKBIA	NM_020529	P25963	Cell signaling / receptor
93	Il-11 receptor chain	IL11RA	NM_004512	Q16542	Cell signaling / receptor
94	Interleukine 1 alpha	IL1A	NM_000575	P01583	Growth factor and cytokines
95	Interleukin 15	IL15	NM_000585	P40933	Growth factor and cytokines
96	Interleukin 6	IL6	NM_000600	P05231	Growth factor and cytokines
97	Interleukin 6 receptor	IL6R	NM_000565	P08887	Cell signaling / receptor
98	Interleukin 11	IL11	NM_000641	P20809	Growth factor and cytokines
99	Mitogen activated protein kinase8	MAPK8	L26318	P45983	Stress response
100	Mitogen activated protein kinase9	MAPK9	U09759	P45984	Stress response
101	Malate dehydrogenase 1	MDH1	NM_005917	P40925	HouseKeeping gene
102	mitogen-activated protein kinase 10	MAPK10	NM_002753	P53779	Stress response
103	monoamine oxidase A	MAOA	NM_000240	P21397	Drug metabolism
104	monoamine oxidase B	MAOB	NM_000898	P27338	Drug metabolism
105	Multidrug resistance 1	ABCB1	NM_000927	P08183	Multidrug Resistance Transporter
106	Phospholipase A2	YWHAZ	M86400	P04054	HouseKeeping gene
107	O-6-methylguanine-DNA methyltransferase	MGMT	NM_002412	P16455	DNA repair / synthesis
108	MnSoD	SOD2	NM_000636	P04179	Oxidative metabolism
109	Multidrug resistance-associated protein 1	ABCC1	AJ003198	P33527	Multidrug Resistance Transporter
110	Multidrug resistance-associated protein 2	ABCC2	NM_000392	Q92887	Multidrug Resistance Transporter
111	Multidrug resistance-associated protein 3	ABCC3	AF085690	O15438	Multidrug Resistance Transporter
112	Multidrug resistance-associated protein 4	ABCC4	AY081219	O15439	Multidrug Resistance Transporter
113	Multidrug resistance-associated protein 5	ABCC5	NM_005688	O15440	Multidrug Resistance Transporter
114	Multidrug resistance-associated protein 6	ABCC6	NM_001171	O95255	Multidrug Resistance Transporter
115	NADPH oxidoreductase	NDUFA5	NM_005000	Q16718	Drug metabolism
116	N-acetyltransferase 1/2	NAT1/2	NM_000662	P18440	Drug metabolism
117	Nuclear factor of kappa B	NFKB1	NM_003998	P19838	Transcription
118	OAT2	SLC22A7	NM_006672	Q9Y694	Organic Anion Transporter
119	OAT3	SLC22A8	NM_004254	O95820	Organic Anion Transporter
120	Organic anion transporter 8	SLC21A8	NM_019844	Q9NPD5	Organic Anion Transporter
121	Organic anion transporter A	SLC21A3	NM_134431	P46721	Organic Anion Transporter
122	Organic anion transporter B	SLC21A9	NM_007256	O94956	Organic Anion Transporter
123	Ribosomal Proteine S9	RPS9	NM_001013	P46781	HouseKeeping gene
124	Organic anion transporter C	SLC21A6	NM_006446	Q9Y6L6	Organic Anion Transporter
125	Organic cation transporter I	SLC22A1	NM_003057	O15245	Organic Cation Transporter
126	Organic cation transporter II	SLC22A2	AB075951	Q9NQB9	Organic Cation Transporter
127	Ornithine decarboxylase1	ODC1	NM_002539	P11926	intermediate metabolism
128	osteonectin	SPARC	NM_003118	P09486	Extracellular matrix
129	Serine Dehydratase	SDS	NM_006843	P20132	HouseKeeping gene
130	Cyclin dependent kinase inhibitor 2A	CDKN2A	L27211	Q8N726	Cell cycle
131	Cyclin dependent kinase inhibitor 1A	CDKN1A	U03106	P38936	Cell cycle
132	Cyclin dependent kinase inhibitor 1B	CDKN1B	NM_004064	P46527	Cell cycle
133	Tumor protein p53	TP53	AF307851	P04637	Cell cycle
134	Proliferating cell nuclear antigen	PCNA	NM_002592	P12004	Cell cycle
135	polysynthetase	ADPRT	J03473	P09874	DNA repair / synthesis
136	PPAR alpha	PPARA	NM_005036	Q07869	Transcription

Table 64 continued:

No.	Gene Name	Official Gene Symbol	Genbank	Swissprot	General Function
137	PPAR delta	PPARD	NM_006238	Q03181	Transcription
138	PPAR gamma	PPARG	NM_015869	P37231	Transcription
139	PEPT1	SLC15A1	NM_005073	P46059	Transport
140	Transferrin receptor	TFRC	NM_003234	P02786	HouseKeeping gene
141	PXR(pregnane X receptor)	NR1I2	NM_003889	O75469	Transcription
142	RXR(retinoic acid receptor)	RARA	NM_000964	P10276	Transcription
143	Senescence marker protein30	RGN	NM_004683	Q15493	Cell cycle
144	SULT1A	SULT1A	AB062428	P50225	Drug metabolism
145	SULT1E1	STE	NM_005420	P49888	Drug metabolism
146	SULT2A1	SULT2A1	NM_003167	Q06520	Drug metabolism
147	Telomerase associated protein 1	TEP1	NM_007110	O00633	Cell cycle
148	Transforming growth factor beta II	TGFBR II	D50683	P37173	Stress response
149	Tumor necrosis factor alpha	TNF	NM_000594	P01375	Growth factor and cytokines
150	UGT 1A	UGT1A	NM_000463	P22309	Drug metabolism
151	UGT2B	UGT2B	NM_001074	P16662	Drug metabolism

Table 65: Differentially expressed genes in human hepatocytes after 24 h induction with Rifampicin based on an analysis using ChiplInspector 1.35 from Genomatix. The normalization was done using a linear total intensity normalization algorithm. For statistical evaluation a significance test was performed at the single probe level. This was done via a standard permutational t-test, similar to SAM Tusher et al., 2001. The resulting list of significantly regulated genes including gene symbol, gene name and the obtained ratio between induced and control cells are given.

No.	Symbol	Gene Name	Ratio
1	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	2.2
2	MT1F	metallothionein 1F	1.8
3	CYP3A43	cytochrome P450, family 3, subfamily A, polypeptide 43	1.7
4	ALAS1	aminolevulinate, delta-, synthase 1	1.6
5	MT1X	metallothionein 1X	1.4
6	CYP3A7	cytochrome P450, family 3, subfamily A, polypeptide 7	1.4
7	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 pg 1	1.4
8	AGXT2L1	alanine-glyoxylate aminotransferase 2-like 1	1.3
9	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	1.3
10	MALAT1	metastasis associated lung adenocarcinoma transcript 1	1.3
11	ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1	1.2
12	MT1E	metallothionein 1E	1.2
13	CYP2C8	cytochrome P450, family 2, subfamily C, polypeptide 8	1.2
14	MT1M	metallothionein 1M	1.2
15	SEC14L4	SEC14-like 4 (<i>S. cerevisiae</i>)	1.1
16	OSTbeta	organic solute transporter beta	1.1
17	MT1G	metallothionein 1G	1.1
18	MT2A	metallothionein 2A	1.1
19	PRODH2	proline dehydrogenase (oxidase) 2	1.1
20	MT1H	metallothionein 1H	1.1
21	GPX2	glutathione peroxidase 2 (gastrointestinal)	1.0
22	CD14	CD14 molecule	1.0
23	AKR1B10	aldo-keto reductase family 1 member B10 (aldose reductase)	1.0
24	MT1P2	metallothionein 1 pseudogene 2	1.0
25	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19	1.0
26	GDF15	growth differentiation factor 15	1.0
27	CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	1.0
28	CYP2A6	cytochrome P450, family 2, subfamily A, polypeptide 6	1.0
29	AKAP12	A kinase (PRKA) anchor protein (gravin) 12	1.0
30	SERPINB9	serpin peptidase inhibitor, clade B (ovalbumin), member 9	1.0
31	TNRC9	trinucleotide repeat containing 9	0.9
32	CA12	carbonic anhydrase XII	0.9
33	AKR1D1	aldo-keto reductase family 1, member D1	0.9
34	MMD	monocyte to macrophage differentiation-associated	0.9
35	UPP1	uridine phosphorylase 1	0.9
36	COL12A1	collagen, type XII, alpha 1	0.9
37	SOX9	SRY (sex determining region Y)-box 9	0.8
38	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F	0.8
39	CYP2C9	cytochrome P450, family 2, subfamily C, polypeptide 9	0.8
40	STK39	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)	0.8
41	PGD	phosphogluconate dehydrogenase	0.7
42	HSPC111	hypothetical protein HSPC111	0.7
43	RCC1	regulator of chromosome condensation 1	0.7

Table 65 continued:

No.	Symbol	Gene Name	Ratio
44	TXNRD1	thioredoxin reductase 1	0.7
45	HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	0.7
46	SQSTM1	sequestosome 1	0.7
47	SLC1A4	solute carrier family 1 member 4	0.7
48	FARSB	phenylalanyl-tRNA synthetase, beta subunit	0.7
49	MTCP1	mature T-cell proliferation 1	0.7
50	CHORDC1	cysteine and histidine-rich domain (CHORD)-containing 1	0.7
51	C5orf24	chromosome 5 open reading frame 24	0.7
52	CCDC47	coiled-coil domain containing 47	0.7
53	AZIN1	antizyme inhibitor 1	0.7
54	CD59	CD59 molecule, complement regulatory protein	0.7
55	F13B	coagulation factor XIII, B polypeptide	0.6
56	DKC1	dyskeratosis congenita 1, dyskerin	0.6
57	HSPH1	heat shock 105kDa/110kDa protein 1	0.6
58	ELOVL6	ELOVL family member 6, elongation of long chain fatty acids	0.6
59	GPIAP1	GPI-anchored membrane protein 1	0.6
60	SLC38A1	solute carrier family 38, member 1	0.6
61	TSPAN3	tetraspanin 3	0.6
62	BID	BH3 interacting domain death agonist	0.6
63	RHOBTB3	Rho-related BTB domain containing 3	-0.6
64	NIT1	nitrilase 1	-0.6
65	PON3	paraoxonase 3	-0.6
66	CIDEB	cell death-inducing DFFA-like effector b	-0.6
67	C5orf4	chromosome 5 open reading frame 4	-0.6
68	DHRS1	dehydrogenase/reductase (SDR family) member 1	-0.6
69	CTSO	cathepsin O	-0.6
70	PBX3	pre-B-cell leukemia homeobox 3	-0.6
71	OBSL1	obscurin-like 1	-0.6
72	ESR1	estrogen receptor 1	-0.6
73	HTR2B	5-hydroxytryptamine (serotonin) receptor 2B	-0.6
74	LIPG	lipase, endothelial	-0.7
75	KLC4	kinesin light chain 4	-0.7
76	IL1RN	interleukin 1 receptor antagonist	-0.7
77	GRAMD1C	GRAM domain containing 1C	-0.7
78	METTL7A	methyltransferase like 7A	-0.7
79	ID2	inhibitor of DNA binding 2, negative helix-loop-helix protein	-0.7
80	SLC43A3	solute carrier family 43, member 3	-0.7
81	CXXC5	CXXC finger 5	-0.7
82	RAB26	RAB26, member RAS oncogene family	-0.7
83	RHOB	ras homolog gene family, member B	-0.7
84	G0S2	G0/G1switch 2	-0.7
85	PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	-0.7
86	APOL3	apolipoprotein L, 3	-0.7
87	MTMR4	myotubularin related protein 4	-0.7
88	SALL1	sal-like 1 (Drosophila)	-0.7
89	BHMT	betaine-homocysteine methyltransferase	-0.7
90	ESPN	espin	-0.7

Table 65 continued:

No.	Symbol	Gene Name	Ratio
91	SORBS2	sorbin and SH3 domain containing 2	-0.7
92	AS3MT	arsenic (+3 oxidation state) methyltransferase	-0.7
93	MXI1	MAX interactor 1	-0.7
94	MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog	-0.7
95	PPP1R3C	protein phosphatase 1, regulatory (inhibitor) subunit 3C	-0.7
96	DEPDC6	DEP domain containing 6	-0.7
97	LECT2	leukocyte cell-derived chemotaxin 2	-0.7
98	HSD17B6	hydroxysteroid (17-beta) dehydrogenase 6 homolog (mouse)	-0.7
99	KCNN2	potassium conductance calcium-activated channel, N, member 2	-0.7
100	SON	SON DNA binding protein	-0.7
101	CTSB	cathepsin B	-0.7
102	SLC39A10	solute carrier family 39 (zinc transporter), member 10	-0.7
103	C6orf142	chromosome 6 open reading frame 142	-0.7
104	NEB	nebulin	-0.7
105	UPB1	ureidopropionase, beta	-0.7
106	SDCBP2	syndecan binding protein (syntenin) 2	-0.8
107	MAN1C1	mannosidase, alpha, class 1C, member 1	-0.8
108	GALT	galactose-1-phosphate uridylyltransferase	-0.8
109	SORL1	sortilin-related receptor, L(DLR class) A repeats-containing	-0.8
110	FYN	FYN oncogene related to SRC, FGR, YES	-0.8
111	LIME1	Lck interacting transmembrane adaptor 1	-0.8
112	HNF4A	hepatocyte nuclear factor 4, alpha	-0.8
113	CYP8B1	cytochrome P450, family 8, subfamily B, polypeptide 1	-0.8
114	C1orf21	chromosome 1 open reading frame 21	-0.8
115	MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B	-0.8
116	FOLH1	folate hydrolase (prostate-specific membrane antigen) 1	-0.8
117	LMO7	LIM domain 7	-0.8
118	OLFML1	olfactomedin-like 1	-0.8
119	CITED2	Cbp/p300-interacting transactivator, with c-terminal domain, 2	-0.8
120	GPR88	G protein-coupled receptor 88	-0.8
121	GLS2	glutaminase 2 (liver, mitochondrial)	-0.8
122	KIAA1305	KIAA1305	-0.8
123	GYS2	glycogen synthase 2 (liver)	-0.8
124	ASPA	aspartoacylase (Canavan disease)	-0.8
125	RXRA	retinoid X receptor, alpha	-0.8
126	KLKB1	kallikrein B, plasma (Fletcher factor) 1	-0.8
127	ATP2B2	ATPase, Ca ⁺⁺ transporting, plasma membrane 2	-0.8
128	IGF1	insulin-like growth factor 1 (somatomedin C)	-0.8
129	FLJ36031	hypothetical protein FLJ36031	-0.8
130	MOGAT2	monoacylglycerol O-acyltransferase 2	-0.9
131	S100A14	S100 calcium binding protein A14	-0.9
132	AFM	afamin	-0.9
133	PKLR	pyruvate kinase, liver and RBC	-0.9
134	HAO2	hydroxyacid oxidase 2 (long chain)	-0.9
135	COL5A2	collagen, type V, alpha 2	-0.9
136	SLC22A7	solute carrier family 22 (organic anion transporter), member 7	-0.9
137	PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)	-0.9

Table 65 continued:

No.	Symbol	Gene Name	Ratio
138	PDK4	pyruvate dehydrogenase kinase, isozyme 4	-0.9
139	IFIT2	interferon-induced protein with tetratricopeptide repeats 2	-0.9
140	SLC6A1	solute carrier family 6 member 1	-0.9
141	RDH16	retinol dehydrogenase 16 (all-trans)	-0.9
142	SLC38A4	solute carrier family 38, member 4	-0.9
143	ADH6	alcohol dehydrogenase 6 (class V)	-1.0
144	SERPINA10	serpin peptidase inhibitor, clade A (alpha-1), member 10	-1.0
145	INHBE	inhibin, beta E	-1.0
146	MAT2A	methionine adenosyltransferase II, alpha	-1.0
147	ADH1A	alcohol dehydrogenase 1A (class I), alpha polypeptide	-1.1
148	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	-1.1
149	CYP2E1	cytochrome P450, family 2, subfamily E, polypeptide 1	-1.1
150	ADH4	alcohol dehydrogenase 4 (class II), pi polypeptide	-1.2
151	NR1I3	nuclear receptor subfamily 1, group I, member 3	-1.2
152	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2	-1.3
153	TAT	tyrosine aminotransferase	-1.3
154	H19	H19, imprinted maternally expressed untranslated mRNA	-1.3
155	CYP7A1	cytochrome P450, family 7, subfamily A, polypeptide 1	-1.4
156	ADH1B	alcohol dehydrogenase IB (class I), beta polypeptide	-1.6

Curriculum Vitae

Oberbrunnweg 25 • 71069 Sindelfingen
Telefon 07031 463927 • 0162 8799874
E-mail Thomas.Reichart@itb.uni-stuttgart.de



THOMAS REICHART

PERSÖNLICHE INFORMATION

Geburtsdatum: 13.11.1976
Geburtsort: Sindelfingen
Familienstand: ledig
Staatsangehörigkeit: Deutsch

SCHULAUSBILDUNG

1987 – 1993	Realschule Hinterweil
1993 – 1996	Technisches Gymnasium Sindelfingen

Abschluss: Allgemeine Hochschulreife, Gesamtnote 1,6

ZIVILDIENST

11/1996 – 10/1997 Pflegezentrum Maichingen des Evangelischen Diakonievereins e.V.

AKADEMISCHE AUSBILDUNG

10/1997 - 03/2004	Studium der Technischen Biologie an der Universität Stuttgart
-------------------	---------------------------------------------------------------

- Studienschwerpunkte: Technische Biochemie, Immunologie sowie Mikrobiologie und Bioverfahrenstechnik und Biochemie
- Praktika in den Bereichen Biochemie, Bioinformatik, Bioverfahrenstechnik, Genetik, Mikrobiologie, Technische Biochemie

Abschluss: Diplom-Biologe (technisch orientiert), sehr gut

PRAKTISCHE TÄTIGKEITEN

09/2002 Studienarbeit am Institut für Mikrobiologie der Universität Stuttgart:
„His-tag Modifizierung und Reinigung einer PHB-Depolymerase aus *Rhodospirillum rubrum*.“

03/2004 Diplomarbeit am Institut für Mikrobiologie der Universität Stuttgart:
„Untersuchungen zur Funktion und Interaktion zwischen PHB-Polymerase, PHB granula-assoziierten Proteinen und ApdA aus *Rhodospirillum rubrum*.“

06/2008 Promotion am Institut für Technische Biochemie der Universität Stuttgart:
„Geneexpressionsanalyse von primären humanen Hepatozyten mittels DNA Mikroarrays als Grundlage für die dynamische Modellierung der menschlichen Leber - ein systembiologischer Ansatz“

Abschluss: Dr. rer. nat., sehr gut