Humboldt-Universität zu Berlin

Lebenswissenschaftliche Fakultät
Institut für Biologie

Bachelorarbeit

Zum Erwerb des akademischen Grades Bachelor of Science

A systematic overview of protein variability in cytochrome P450 and UDP-glucuronosyltransferase enzymes in the human liver

Ein systematischer Überblick über die Proteinvariabilität von Cytochrom P450 und UDP-Glucuronosyltransferase-Enzymen in der menschlichen Leber

vorgelegt von
Afruja Hossain
Matrikelnummer: 576504
E-Mail: hossaina@hu-berlin.de
Geburtsdatum und -ort: 24.07.1996, Comilla, Bangladesch
angefertigt in der Arbeitsgruppe Systems Medicine of the Liver
am Institut für Theoretische Biologie
Berlin, im Dezember 2023
Erstprüfer: **Dr. Matthias König**
Zweitprüfer:

Contents

Abstract

English

Cytochrome P450 (CYP450) and UDP-glucuronosyltransferase (UGT) enzymes play key roles in the metabolism of a wide variety of endogenous and exogenous compounds, including drugs and toxins. CYP450 enzymes are involved in phase I metabolic reactions that typically introduce or unmask functional groups on molecules, often leading to their activation or inactivation. These reactions include hydroxylation, oxidation, and reduction. UGT enzymes, on the other hand, are essential for phase II metabolism, where they catalyze the conjugation of glucuronic acid to various substrates, making the compounds more water soluble and therefore easier to excrete. Both CYP450 and UGT are critical in determining drug pharmacokinetics, drug efficacy, drug interactions and potential toxicity.

A variety of CYP450 and UGT isoforms exist in the human liver, and different isoforms are responsible for detoxifying different substances. For example, CYP1A2 is involved in the metabolism of caffeine, CYP2D6 in dextromethorphan, CYP2E1 in chlorzoxazone, and CYP3A4 in midazolam. There are large inter-individual differences in the metabolism of drugs by the different CYP450 and UGT isoforms. Important factors for these differences are differences in activity, e.g. due to genetic polymorphisms, and the amount of protein of the respective enzyme in the liver.

The aim of this study was to provide a systematic overview of the variability in protein levels of CYP450 and UGT isoforms in the human liver. To this end, a manually curated database of protein abundance information was established and used to answer the following questions: (i) What are the protein distributions of the different CYP450 and UGT isoforms? (ii) What factors affect these protein levels (age, sex, body mass index, smoking, alcohol consumption)? (iii) How do CYP and UGT isoforms correlate among each other?

Key results include: (i) A free and open database of CYP and UGT protein levels in the human liver has been established. (ii) CYP and UGT isoforms show large inter-individual variability in protein levels. (iii) Different CYP and UGT isoforms have different protein levels. (iv) Factors such as age, smoking and alcohol consumption may affect the protein levels of CYP and UGT enzymes. (v) There are significant correlations in protein levels between CYP isoforms and between UGT isoforms.

This work has established an important resource for the study of variability in drug detoxification due to variability in protein levels in the human liver.

German

Die Enzyme Cytochrom P450 (CYP450) und UDP-Glucuronosyltransferase (UGT) spielen eine Schlüsselrolle beim Stoffwechsel einer Vielzahl von endogenen und exogenen Verbindungen, einschließlich Medikamenten und Toxinen. CYP450-Enzyme sind an Stoffwechselreaktionen der Phase I beteiligt, bei denen typischerweise funktionelle Gruppen in Moleküle eingeführt oder demaskiert werden, was häufig zu deren Aktivierung oder Inaktivierung führt. Zu diesen Reaktionen gehören Hydroxylierung, Oxidation und Reduktion. UGT-Enzyme hingegen sind für den Phase-II-Stoffwechsel unerlässlich, wo sie die Konjugation von Glucuronsäure an verschiedene Substrate katalysieren, wodurch die Verbindungen wasserlöslicher werden und daher leichter ausgeschieden werden können. Sowohl CYP450 als auch UGT sind von entscheidender Bedeutung für die Pharmakokinetik von Arzneimitteln, die Wirksamkeit von Arzneimitteln, Arzneimittelwechselwirkungen und potenzielle Toxizität.

In der menschlichen Leber gibt es eine Vielzahl von CYP450- und UGT-Isoformen, und verschiedene Isoformen sind für die Entgiftung unterschiedlicher Substanzen zuständig. So ist beispielsweise CYP1A2 am Stoffwechsel von Koffein beteiligt, CYP2D6 an Dextromethorphan, CYP2E1 an Chlorzoxazon und CYP3A4 an Midazolam. Es gibt große interindividuelle Unterschiede beim Metabolismus von Arzneimitteln durch die verschiedenen CYP450- und UGT-Isoformen. Wichtige Faktoren für diese Unterschiede sind Aktivitätsunterschiede, z. B. aufgrund genetischer Polymorphismen, und die Menge an Protein des jeweiligen Enzyms in der Leber.

Ziel dieser Studie war es, einen systematischen Überblick über die Variabilität der Proteinmenge von CYP450- und UGT-Isoformen in der menschlichen Leber zu geben. Zu diesem Zweck wurde eine manuell kuratierte Datenbank mit Informationen zur Proteinhäufigkeit erstellt und zur Beantwortung der folgenden Fragen verwendet: (i) Wie sind die Proteinverteilungen der verschiedenen CYP450- und UGT-Isoformen? (ii) Welche Faktoren beeinflussen diese Proteinmengen (Alter, Geschlecht, Body-Mass-Index, Rauchen, Alkoholkonsum)? (iii) Wie korrelieren die CYP- und UGT-Isoformen miteinander?

Die wichtigsten Ergebnisse sind: (i) Es wurde eine freie und offene Datenbank für CYP- und UGT-Proteinspiegel in der menschlichen Leber eingerichtet. (ii) CYP- und UGT-Isoformen weisen eine große interindividuelle Variabilität der Proteinkonzentrationen auf. (iii) Verschiedene CYP- und UGT-Isoformen weisen unterschiedliche Proteingehalte auf. (iv) Faktoren wie Alter, Rauchen und Alkoholkonsum können die Proteinkonzentrationen der CYP- und UGT-Enzyme beeinflussen. (v) Es gibt signifikante Korrelationen der Proteingehalte zwischen CYP-Isoformen und zwischen UGT-Isoformen.

Mit dieser Arbeit wurde eine wichtige Ressource für die Untersuchung der Variabilität bei der Drogenentgiftung aufgrund der Variabilität der Proteingehalte in der menschlichen Leber geschaffen.

1 Introduction

1.1 Drug detoxification

Drug detoxification is a critical process that removes unwanted substances from the body. It involves biotransformations that convert both endogenous and exogenous compounds into more water-soluble products to facilitate their elimination. Drug metabolism occurs primarily in the liver, but metabolizing enzymes can be found throughout the body [35]. The drug detoxification process can be divided into three stages: modification, conjugation, and excretion. The liver is primarily responsible for the metabolism of drugs during phases I and II. As a result of this metabolic process, compounds often become more water-attractive (hydrophilic), allowing for easy secretion by the kidneys [4].

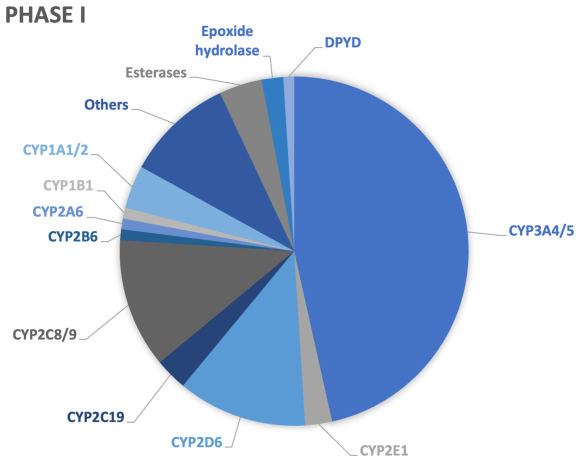


Figure 1: **Enzymes involved in phase I drug detoxification**. CYP450 enzymes are involved in phase I metabolic reactions that typically introduce or unmask functional groups on molecules, often resulting in their activation or inactivation. Such reactions include hydroxylation, oxidation, and reduction. The contribution of different enzymes to the phase I metabolism of drugs is indicated by the size of each section of the diagram. CYP, cytochrome P450; DPD, dihydropyrimidine dehydrogenase; GST, glutathione S-transferase; NAT, N-acetyltransferase; STs, sulfotransferases; TPMT, thiopurine methyltransferase; UGTs, uridine 59-triphosphate glucuronosyltransferases. Adapted from [14].

In phase I, several enzymes, especially the cytochrome P-450-dependent mixed-function oxidase system, facilitate the addition of reactive and polar groups (Fig. 1). Hydroxylation is one of the most common modifications during this phase. Phase I reactions, also known as non-synthetic reactions, can take place through processes such as oxidation, reduction, hydrolysis, cyclization, decyclization, and either the introduction of oxygen or the removal of hydrogen. These reactions are mainly carried out by mixed-function oxidases, primarily in the liver. The typical oxidative reactions in this phase involve a cytochrome P450 monooxygenase (commonly abbreviated as CYP), NADPH, and oxygen [21].

Phase II reactions involve the conjugation of activated xenobiotic metabolites with charged entities such as glutathione (GSH), sulfate, glycine, or glucuronic acid (Fig. 2). The result of these conjugation reactions is the formation of higher molecular weight products that are generally less active than their precursors. This is a notable difference from phase I reactions, which often produce active metabolites.

After the phase II reactions, the xenobiotic conjugates may undergo further metabolic processes (Fig. 3). A classic example is the conversion of glutathione conjugates into acetylcysteine (mercapturic acid) conjugates [7]. During this transition, enzymes such as gamma-glutamyl transpeptidase and dipeptidases act to remove the gamma-glutamate and glycine residues from the glutathione core structure. The culmination of this process is the acetylation of the cysteine residue present in the conjugate. In the subsequent Phase III of metabolism, these conjugates and their respective metabolites are prepared for elimination [23]. The anionic group present in them serves as

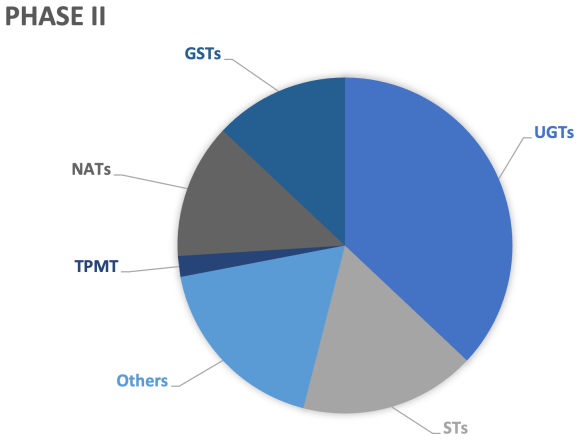


Figure 2: **Enzymes involved in phase II of drug detoxification**. UGT enzymes, on the other hand, are essential for phase II metabolism, where they catalyze the conjugation of glucuronic acid to various substrates, making the compounds more water soluble and therefore easier to excrete. The percentage contribution different enzymes to the phase II metabolism of drugs is indicated by the size of each section of the diagram. GST, glutathione S-transferase; NAT, N-acetyltransferase; STs, sulfotransferases; TPMT, thiopurine methyltransferase; UGTs, uridine-triphosphate glucuronosyltransferases. Adapted from [14].

a marker, ensuring their affinity to various multidrug resistance protein (MRP) membrane transporters, facilitating their removal [26, 9].

Cytochrome P450 (CYP450) and UDP-glucuronosyltransferase (UGT) are key enzymes responsible for the metabolism of a wide range of endogenous and exogenous compounds, including drugs and toxins. CYP450 enzymes are primarily involved in Phase I metabolic reactions. These reactions typically involve the introduction or exposure of functional groups on molecules, leading to either their activation or inactivation. Examples of such reactions include hydroxylation, oxidation, and reduction.

Conversely, UGT enzymes are integral to phase II metabolism (Fig. 3). Their primary function in this phase is to catalyze the attachment of glucuronic acid to various substrates. This conjugation increases the water solubility of the compounds, thereby facilitating their elimination.

Both CYP450 and UGT significantly influence drug pharmacokinetics, ensuring optimal drug efficacy, modulating drug-drug interactions, and monitoring potential toxicity.

1.2 Cytochrome P450 (CYP)

Cytochrome P450 (CYP) proteins are the best-known drug-metabolizing enzymes expressed primarily in the liver. CYPs, especially those involved in drug metabolism, play an essential role in maintaining overall human health [44].

A variety of CYP450 isoforms exist in the human liver, and different isoforms are responsible for detoxifying different substances. Some CYPs metabolize only one (or very few) substrates, such as CYP19 (aromatase). For example, CYP1A2 is involved in the metabolism of caffeine, CYP2C19 in omeprazole, CYP2E1 in chlorzoxazone, or CYP3A4 in midazolam (Fig. 4).

Human cytochrome P450 (CYP) enzymes are membrane-bound hemoproteins expressed primarily in the liver. CYP450s are known as the major class of enzymes that chemically transform drugs into their water-soluble byproducts to aid in elimination by the kidneys and/or liver [4].

In addition to the liver, they are expressed in many other tissues, including the intestinal mucosa (probably the most important extrahepatic tissue for drug metabolism), brain, kidney, lung, and skin, and are important for hormone synthesis and breakdown (including synthesis and metabolism of estrogen and testosterone), cholesterol synthesis, and vitamin D metabolism.

Cytochrome P450 (CYP) is one of the largest protein families consisting of 57 putative functional genes whose products perform a variety of biological oxidations and reductions in drug, xenobiotic, and endogenous metabolism. The CYP3A subfamily consists of at least three isoforms

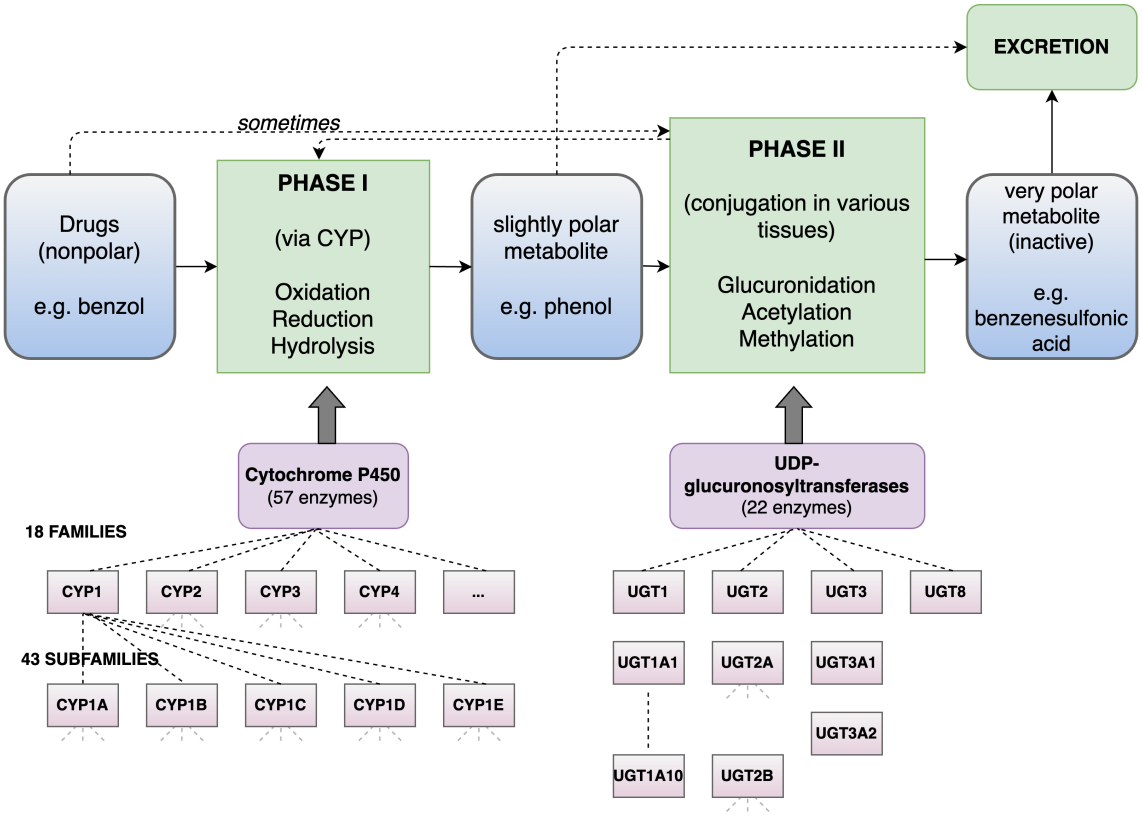


Figure 3: **Schematic overview of phase I and phase II of drug metabolism.** The process of drug metabolism can be categorized into two key phases according to their chemical characteristics. Phase I metabolism involves the oxidation, reduction or hydrolysis of the drug, resulting in metabolites that may be inactive, active, or toxic. This process is carried out primarily by a family of enzymes known as cytochrome P450 (CYP) enzymes. Phase II metabolism involves the conjugation of the modified drug with another molecule, such as glucuronic acid, sulfate, or amino acids. This process increases the water solubility of the drug and helps it to excrete easier. Phase II reactions are commonly catalyzed by various enzymes, such as UDP-glucuronosyltransferases (UGTs), sulfotransferases, and glutathione S-transferases (GSTs). Adapted from [37].

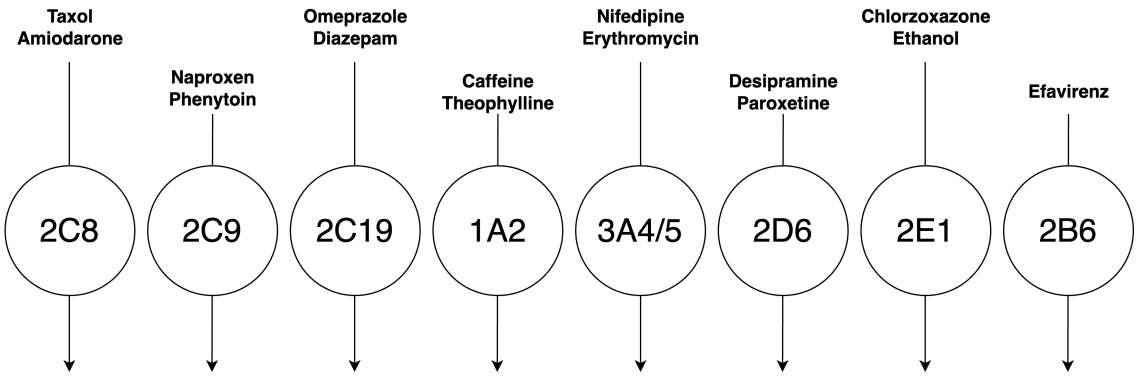


Figure 4: **Examples of different substrate specificities of human cytochrome P450 enzymes**. Representative substrates are shown above each enzyme. Adapted from [12].

and is the most abundant group of CYP enzymes in the liver: CYP3A4, 3A5, and 3A7. This subfamily is the major metabolizer of a number of drugs [13]. CYP3A4 is the most widely expressed CYP, accounting for approximately 30 to 40 percent of the total CYP content in adult human liver

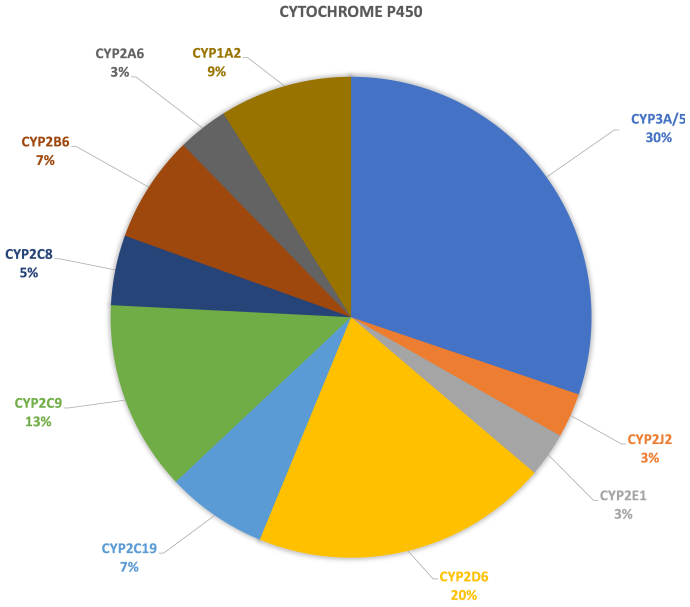


Figure 5: **Overview of clinically relevant drug-metabolizing cytochrome P450 isoforms**. There are 57 sequenced P450 genes in humans, with 18 families and 43 subfamilies of P450s. Of the 57 genes, 10-13 are involved in the majority of P450 drug metabolism. These include CYP3A4/5, CYP2C9, CYP2C8, CYP2C19, CYP2D6, CYP2B6, CYP2E1, CYP2A6, and CYP1A2, with CYP3A4 and CYP2D6 contributing to over 50% of P450-related drug metabolism. Adapted from [43].

and small intestine. CYP3A5 is 83 percent homologous to CYP3A4 and has a much lower level of expression in the liver than CYP3A4, but is the major CYP3A isoform in the kidney. CYP3A7 is the major CYP isoform detected in human embryonic, fetal and neonatal liver, but is also detected in adult liver, but at much lower levels than CYP3A4 [13].

1.3 UDP-glucuronosyltransferase (UGT)

Uridine 5'-diphospho-glucuronosyltransferase (UDP-glucuronosyltransferase, UGT) is a microsomal glycosyltransferase commonly involved in the metabolism of drugs, xenobiotics, bilirubin, androgens, estrogens, mineralocorticoids, glucocorticoids, fatty acid derivatives, retinoids, and bile acids.

UGTs are membrane-bound conjugating enzymes that catalyze the transfer of the glucuronic acid group of uridine diphosphoglucuronic acid (UDP-GlcA) to the functional group (e.g. hydroxyl, carboxyl, amino, sulfur) of a specific substrate. Human UDP-glucuronosyltransferases (UGTs) are a superfamily of 22 proteins divided into 5 families and 6 subfamilies based on sequence identity. The UGT1A and 2B subfamilies play an essential role in terminating the biological effects and enhancing the renal elimination of nonpolar (lipophilic) drugs of all therapeutic classes. The primary function of these enzymes is to catalyze the covalent attachment of glucuronic acid, generated from the cofactor UDP-glucuronic acid, to a substrate with an appropriate acceptor functional group. This process is called glucuronidation. While the liver is the major detoxification organ and as such contains the greatest abundance and diversity of UGTs, these enzymes also have significant but variable expression outside the liver [32]. The phase II glucuronidation reaction consists of the transfer of the glucuronosyl group from uridine-5'-diphospho-glucuronic acid (UDPGA) to substrate molecules containing oxygen, nitrogen, sulfur, or carboxyl functional groups. The resulting glucuronide is more polar (e.g., hydrophilic) and easier to excrete than the substrate molecule. The solubility of the product in the blood is increased, allowing it to be eliminated from the body by the kidneys.

In addition to liver, UGT1A and UGT2B enzymes are expressed in a number of other tissues, including kidney, small intestine, colon, stomach, lung, epithelium, ovary, testis, and prostate [32].

1.4 Question, scope and hypotheses

Inter-individual variability in drug metabolism and liver function is an important factor in determining how different people respond to drugs. A deeper understanding of protein variability and the factors that influence it is fundamental to understanding this phenomenon. The specific protein levels of the CYP450 and UGT isoforms involved in the metabolic process are central to the differences in drug elimination rates between individuals. Current reports on CYP450 and UGT protein levels lack comprehensiveness, and individual data are not easily accessible for further analysis or reuse [2]. In particular, there is currently no open, freely available database or resource detailing CYP450 and UGT protein levels in the human liver.

The aim of this study was to provide a systematic overview of the variability in protein levels of the CYP450 and UGT isoforms in the human liver. Specifically, this work attempts to answer the following questions

- (i) What are the protein distributions of the different CYP450 and UGT isoforms?
- (ii) What factors affect these protein levels (age, sex, body mass index, smoking, alcohol consumption, ethnicity)?

The main objective was to better understand the variability of important drug-metabolizing enzymes in the human liver and to provide a open and free resource of this information for further research.

2 Methods

This chapter describes the methods and resources used to create the database of CYP and UGT proteins in human liver. Several steps were used to provide a systematic overview of protein variability, starting with systematic literature search (Sec. 2.1), followed by data curation (Sec. 2.2), data analysis (Sec. 2.3), analysis of factors affecting protein levels (Sec. 2.4), and correlation analysis (Sec. 2.5).

2.1 Systematic literature research

To establish a database of cytochrome P450 and UDP-glucuronosyltransferase levels in human liver, a systematic literature search was performed.

The PubMed database was searched using an appropriate combination of keywords (`drug metabolism`, `cytochrome P450`, `UDP-glucuronosyltransferase`, `liver OR hepatic`, `human`, `abundance`). In addition to `abundance`, other keywords such as `measurement`, `quantification`, `concentration`, `quantity` were used instead of `abundance` to broaden the search scope.

The main inclusion criteria were that the study was conducted in humans and that the study reported protein abundance data in individual subjects or averaged data. No exclusions based on subject gender, age group, ethnic background, medical history, or other variables were made in order to obtain a broader range of results.

To identify additional literature, references cited in the primary corpus were also examined. Selected studies that cited metadata and individual liver microsomal samples in which protein abundance was measured were selected. To ensure that the data used in the analysis were not duplicated, the sources of the data were identified.

Zotero was used to manage the collected publications. To organize the different references with different enzyme data, groups were created such as `CYP450` and `UGT`. Within the groups, tags were used to organize and select the references (e.g. `priority high` for the more important references). After a thorough review of the literature, the relevant papers were selected for data curation.

2.2 Data curation

A standardized data presentation format was established, and relevant data were manually curated from the literature into this format. The data consisted of both mean and individual protein levels of cytochrome P450 isoforms and UDP-glucuronosyltransferase isoforms and corresponding subject information. Data from tables and figures were extracted and curated into a standardized xlsx spreadsheet file format. The `Groups` sheet coded all reported group information such as number of subjects, health status, age, sex, weight, BMI, smoking status, alcohol consumption, ethnicity, cause of death, medical history, and medications. The sheet `Individuals` encoded the corresponding information for individuals. The remaining sheets encoded protein abundance from tables or figures, and PlotDigitizer was used as a tool to extract numerical data from figures and graphs.

Curation followed established curation workflows for the open pharmacokinetics database PK-DB (https://pk-db.com [19, 18]). Validation scripts were established to ensure the high quality of the data. A second curator checked all curated data for consistency.

2.3 Data analysis

The data from each study were combined into a single data set. All analyses were performed on the individual subject data (i.e., mean data for groups are part of the database but are not used in any of the analyses). The data were filtered for analysis. First, protein isoforms with <10 data points were removed as there was not enough data for analysis. Second, this project focused on the human liver, so data from other subjects were filtered as well as data from other tissues such as intestine. Data sets for both liver and isolated hepatocytes were included in the analysis because the data were highly consistent.

Heterogeneity analysis was performed as reported in [2]. Weighted means (WX) and weighted coefficient of variations (WCV) of enzyme abundances from the combined studies were calculated according to [5].

$$W\overline{X} = \frac{\sum_{j=1}^{J} n_j \cdot \overline{X}_j}{\sum_{j=1}^{J} n_j}$$

$$WCV = \frac{\sum_{j=1}^{J} n_j \cdot CV_j}{\sum_{j=1}^{J} n_j}$$

Subscript j indicates the study, n_j indicates the number of samples in study j, and $\overline{X}_j$ indicates the mean abundance of a particular enzyme in study j.

To assess the heterogeneity between the means and coefficients of variation of individual studies and the overall mean and variability of the pooled data were used:

$$VarW\overline{X} = \frac{\sum_{j=1}^{J} w_j \cdot \overline{X}_j}{\sum_{j=1}^{J} w_j}$$

$$w_j = \frac{1}{(sd_j)^2}$$

$$Q = \sum_{j=1}^{J} \left(w_j \cdot (\overline{X}_j - VarW\overline{X})^2 \right)$$

where w_j is the weight of study j, expressed in terms of variance (calculated as the inverse of the square of the standard deviation, sd_j), $VarWX$ is the variance of the weighted mean of the data from all studies. Q is the coefficient of heterogeneity of the pooled data (Cochran's Q test [8], expressed as the collective weighted squared differences between the mean of each study and the variance of the weighted mean. A higher value of Q indicates greater heterogeneity. The I^2 can be used to assess the degree of heterogeneity [22]. This index provides a percentage of total heterogeneity, which is interpreted as about 0 percent - no heterogeneity, about 25 percent - low heterogeneity, about 50 percent - moderate heterogeneity, and about 75 percent - high heterogeneity.

$$I^2 = 100 \cdot \frac{Q - (k-1)}{Q}$$

where I^2 is the index of heterogeneity, Q is Cochran's coefficient of heterogeneity, and $(k-1)$ is the number of degrees of freedom, defined as the number of studies, k, minus one. If I^2 is negative, it is set to zero.

2.4 Stratification analysis

An analysis of the factors affecting the protein levels of each isoform was performed, examining the effect of age, sex, body mass index (BMI), smoking, and alcohol consumption. For each protein isoform, individual data were stratified by sex (male or female), smoking (yes or no), alcohol consumption (yes or no), age (middle aged or elderly), BMI (normal weight, overweight, obese). The age groups were defined as middle-aged (age range 35-65 years) and elderly ($>$ 65 years). BMI was divided into three groups with normal weight as 18.5-25 bmi, overweight 25-30 bmi and obese with a bmi $>$ 30. A stratification analysis based on ethnicity was not performed because the majority of the data were from Caucasians and the other ethnicities were underrepresented. Preliminary analysis showed insufficient data for analysis.

Significant differences were tested using two-sample Mann-Whitney U tests in the case of two groups (alcohol, age, smoking, sex) and the Kruskal-Wallis test in the case of multiple groups (bmi), the nonparametric equivalent of one-way ANOVA. Importantly, the Kruskal-Wallis test does not assume normality in the data and is much less sensitive to outliers than one-way ANOVA.

Stratification analysis was performed separately for CYP and UGT data. Correction for multiple testing was performed with a family-wise error rate (FWER) of 0.05 using the Benjamini/Hochberg (BH, non-negative) false discovery rate (FDR) method. All analyses were performed in Python version 3.10.13 using the `science.stats` and `statsmodels` packages. Significance levels after correction for multiple testing were considered significant at the following levels: *0.05, **0.01, and ***0.001.

2.5 Correlation analysis

The correlation matrix was calculated using Pearson's correlation. Significance levels are *0.05, **0.01, and ***0.001, with p-values adjusted for multiple testing using Benjamini and Hochberg. Analysis was performed with R version 4.2.1 and the `corrplot` package. To visualize the symmetric correlation matrix, only the lower triangular matrix is shown with positive correlations in blue and negative correlations in red. The area of the circle is proportional to the correlation coefficient.

Volcano plots were created in Python with cutoffs for showing data with labels being p-value <0.05 and absolute correlation coefficient >0.5.

2.6 Web application

The web application was written as a streamlit application in Python available from https://protein-distribution.streamlit.app/. The data can be filtered by species and protein class. For individual isoforms, the protein abundance data is provided as a CSV for download. Summary plots are generated using matplotlib. All data is available as CC-BY citing this thesis.

3 Results

In this thesis, a protein abundance database for cytochrome P450 (CYP) and UDP-glucuronosyltransferase (UGT) isoforms in human liver was established. This chapter is divided into two main sections. The first presents the results for CYP (Sec. 3.1), followed by the results for UGT (Sec. 3.2). The protein abundance database is available via a web application (Sec. 3.3).

3.1 Cytochrome P450 (CYP)

3.1.1 Protein data (CYP)

An overview of studies containing CYP isoform protein abundances is provided in Fig. 6. In total, 18 studies reported abundance data for 35 CYP isoforms.

Couto2019 [10], Vildhede2015 [40] and Wegler2022 [42] reported abundance data for many isoforms compared to studies such as Asai1996 [6], Inoue2000 [24] and Snwader2000 [34] with only 1-2 different protein isoforms. Some CYP isoforms such as 1A2, 2B6, 2C9, 2E1 were reported in many studies, while 1A1, 27A1, 2B, 2D7, 2U1, 2W1, 4F3B, and 7A1 were reported in only a few studies.

The resulting protein abundances of the CYP isoforms are shown in Fig. 7 and Fig. 8.

Protein	Achour2014	Achour2017	Asai1996	Couto2019	Gao2016	Groer2014	Inoue2000	Kawakami2011	Li2015	Michaels2014	Ohtsuki2012	Olesen2000	Seibert2009	Snawder2000	Vasilogianni2022	Vildhede2015	Wang2015	Wegler2022
CYP2E1	3|	1|7	|14	1|23	|75	1|	2|	2|10	1|		1|	1|3	|4	|39	6|50	|39		|61
CYP1A2	3|	1|5	|14	1|23	|50	1|		2|10	1|		1|	1|3	|3		6|48	|39	1|	|57
CYP3A5	3|	1|4		1|23	|56	1|		2|10	1|	|31	1|				4|27	|39	1|	|91
CYP2C9	3|	1|7		1|23	|89	3|		2|10	1|		1|	1|3			6|46	|40	1|	|91
CYP3A4	3|	1|7		1|23	|57	3|		2|9	1|	|30	1|				6|50	|36	1|	|91
CYP2C19	3|	1|4		1|19	|40	3|		2|10	1|		1|	1|3			6|45	|39	1|	|88
CYP2B6	3|	1|2		1|23	|56	1|		2|10	1|		1|				5|22	|40	1|	|58
CYP2C8	3|	1|8		1|23	|74	1|		2|10	1|		1|				6|41	|40	1|	|59
CYP2D6	3|	1|6		1|16	|64	1|		2|10	1|		1|	1|2			3|15	|38	1|	
CYP2A6	3|	1|7		|23	|82			2|10	1|		1|				6|47	|40	1|	|61
CYP3A7	3|	1|2		1|15								1|				6|19	|34		|44
CYP4F2	3|			1|23						|31							6|48	|38		|91
CYP2C18	3|			1|22													5|30	|37		|89
CYP2J2	3|			1|17													5|38	|37		|91
CYP3A43	3|	1|8		|3				2|10			1|								
CYP4A11				|21							1|							|40		|54
CYP2A7				|6														|1		|26
CYP2A13				1|21														|2		|48
CYP20A1				|23														|19		|91
CYP8B1				|23														|20		|57
CYP1A1				|2														|17		|88
CYP2S1				|7														|11		|91
CYP7B1				|19														|17		|74
CYP39A1				|13														|18		|54
CYP4A22				|3														|11		|54
CYP4F11				1|23														|19		|84
CYP4F12				1|23														|19		|91
CYP4F22				|1														|14		|46
CYP4F3				|22														|20		|75
CYP4V2				|22														|19		|77
CYP51A1				|23														|19		|90
CYP2U1																		|6		|82
CYP4F8				|8																|23
CYP27A1																		|19		|91
CYP7A1																		|18		|23
CYP2W1																				|19

Figure 6: **Studies for cytochrome P450 analysis.** For all studies reporting CYP data, the number of group and individual data points are listed for the different proteins (groups—individuals). CYP data were manually curated from [2, 1, 6, 10, 15, 17, 24, 25, 27, 28, 29, 30, 33, 34, 39, 40, 41, 42].

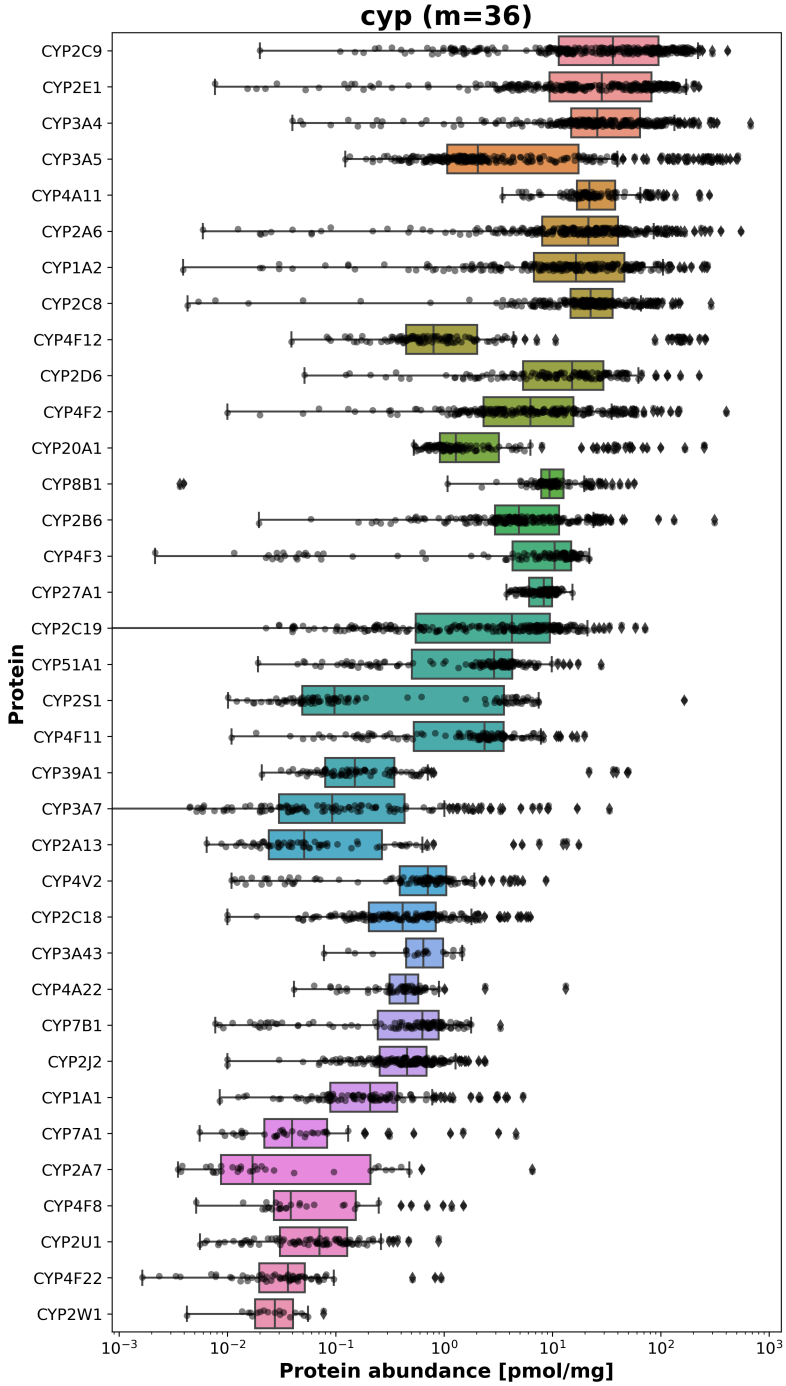


Figure 7: **Protein abundance of cytochrome P450 isoforms.** The box plot and individual scatter plots are shown for each isoform. Each point corresponds to a single measurement of protein amount in a human liver sample from a single subject.

protein	mean	mode	cv	wcv	min	max	n	unit	Q	I^2	df	heterogeneity
CYP2C9	60.19	1.51	1.04	0.68	0.02	413.21	309	pmol/mg	14.48	51.66	7	medium
CYP2E1	48.47	4.06	0.97	0.52	0.01	225.58	325	pmol/mg	61.03	83.61	10	high
CYP3A4	48.36	0.04	1.31	0.82	0.04	675.62	303	pmol/mg	4.60	0.00	7	none
CYP3A5	39.21	0.35	2.37	1.11	0.12	514.54	281	pmol/mg	21.44	67.34	7	medium
CYP4A11	38.57	3.45	1.15	0.44	3.45	282.65	115	pmol/mg	3.15	36.42	2	low
CYP2A6	35.95	0.06	1.53	0.74	0.01	549.68	270	pmol/mg	4.38	0.00	6	none
CYP1A2	32.06	32.50	1.34	0.81	0.00	270.78	252	pmol/mg	39.55	77.25	9	high
CYP2C8	31.32	0.00	1.00	0.56	0.00	292.35	255	pmol/mg	6.32	5.05	6	none
CYP4F12	24.45	0.04	2.43	0.78	0.04	260.67	133	pmol/mg	4.21	52.49	2	medium
CYP2D6	22.12	8.27	1.28	0.80	0.05	227.60	151	pmol/mg	6.46	7.17	6	none
CYP4F2	16.10	0.32	2.10	0.76	0.01	402.20	231	pmol/mg	4.98	19.68	4	none
CYP20A1	12.94	0.53	2.83	0.54	0.53	252.51	133	pmol/mg	2.86	29.99	2	low
CYP8B1	12.54	0.00	0.75	0.36	0.00	56.91	100	pmol/mg	2.76	27.53	2	low
CYP2B6	11.11	0.02	2.25	1.04	0.02	313.09	211	pmol/mg	4,711.79	99.87	6	high
CYP4F3	9.47	0.00	0.66	0.59	0.00	22.00	117	pmol/mg	0.24	0.00	2	none
CYP27A1	8.11	3.78	0.29	0.28	3.78	15.35	110	pmol/mg	0.06	0.00	1	none
CYP2C19	6.27	0.04	1.32	1.11	0.00	71.81	248	pmol/mg	9.37	25.26	7	low
CYP51A1	3.30	0.02	1.08	0.79	0.02	28.13	132	pmol/mg	0.69	0.00	2	none
CYP2S1	3.12	0.01	5.06	1.31	0.01	165.16	109	pmol/mg	0.76	0.00	2	none
CYP4F11	3.02	0.01	1.11	0.69	0.01	19.86	126	pmol/mg	2.01	0.64	2	none
CYP39A1	2.52	0.02	3.78	0.77	0.02	50.26	85	pmol/mg	3.20	37.56	2	low
CYP3A7	1.14	0.04	3.33	1.91	0.00	33.73	114	pmol/mg	2,034.46	99.80	4	high
CYP2A13	0.98	0.01	3.18	1.10	0.01	17.52	71	pmol/mg	0.56	0.00	2	none
CYP4V2	0.96	0.68	1.25	0.76	0.01	8.73	118	pmol/mg	1.21	0.00	2	none
CYP2C18	0.80	0.74	1.40	1.41	0.01	6.27	178	pmol/mg	1.15	0.00	3	none
CYP3A43	0.68	0.08	0.59	0.30	0.08	1.47	21	pmol/mg	12.89	84.48	2	high
CYP4A22	0.65	0.04	2.43	0.48	0.04	13.28	68	pmol/mg	2.02	1.11	2	none
CYP7B1	0.63	0.01	0.81	0.73	0.01	3.32	110	pmol/mg	0.21	0.00	2	none
CYP2J2	0.54	0.11	0.74	0.58	0.01	2.40	183	pmol/mg	2.24	0.00	3	none
CYP1A1	0.49	0.01	1.79	1.54	0.01	5.36	107	pmol/mg	0.96	0.00	2	none
CYP7A1	0.32	0.01	2.71	1.28	0.01	4.62	41	pmol/mg	0.29	0.00	1	none
CYP2A7	0.29	0.00	3.77	1.74	0.00	6.51	33	pmol/mg		0.00	2	none
CYP4F8	0.22	0.01	1.72	0.77	0.01	1.51	31	pmol/mg	2.19	54.40	1	medium
CYP2U1	0.10	0.01	1.18	1.10	0.01	0.89	88	pmol/mg	0.42	0.00	1	none
CYP4F22	0.07	0.00	2.28	0.87	0.00	0.94	61	pmol/mg		0.00	2	none
CYP2W1	0.03	0.00	0.56	-	0.00	0.08	19	pmol/mg	-	-	0	-

Figure 8: **Protein abundance and heterogeneity of cytochrome P450 isoforms.** Weighted means, mode, coefficient of variation (cv), weighted coefficient of variation (wcv), minimum value (min), maximum value (max), number of subjects (n), unit, Cochran's heterogeneity coefficent (Q), Higgen's and Thompson's heterogeneity index (I^2), degree of freedom (df) and heterogeneity analysis of hepatic cytochrome p450 enzyme abundance data.

Examples of protein levels (CYP) Two examples are provided with protein abundances for CYP2E1 shown in Fig. 9 and CYP1A2 in Fig. 10.

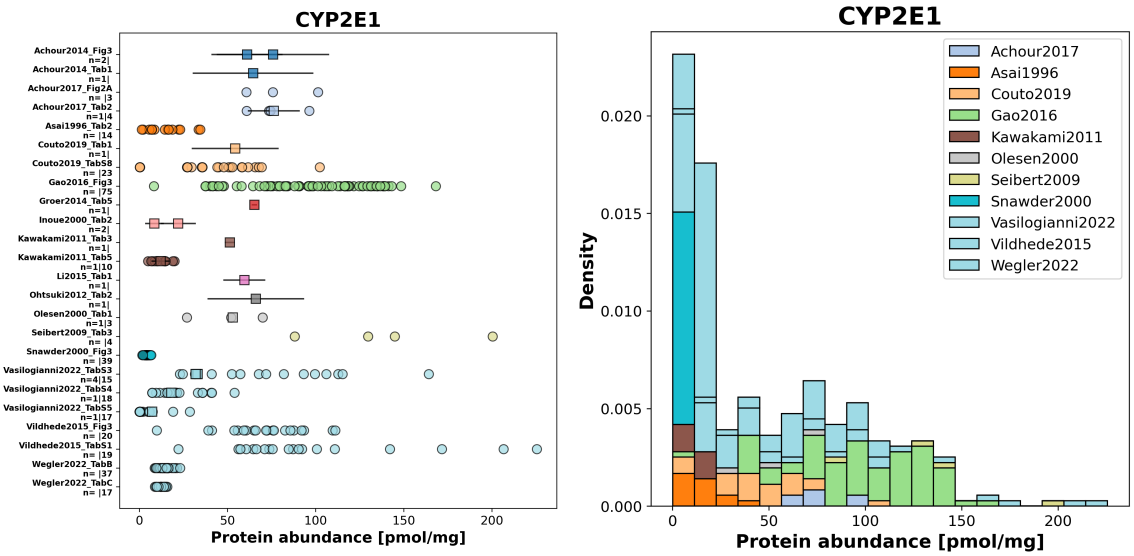


Figure 9: **Protein abundance for CYP2E1.** A) Individual and mean ± SD protein abundance CYP2E1. B) Histogram of individual protein abundance CYP2E1. Data from [2, 1, 6, 10, 15, 17, 24, 25, 27, 29, 30, 33, 34, 39, 40, 42].

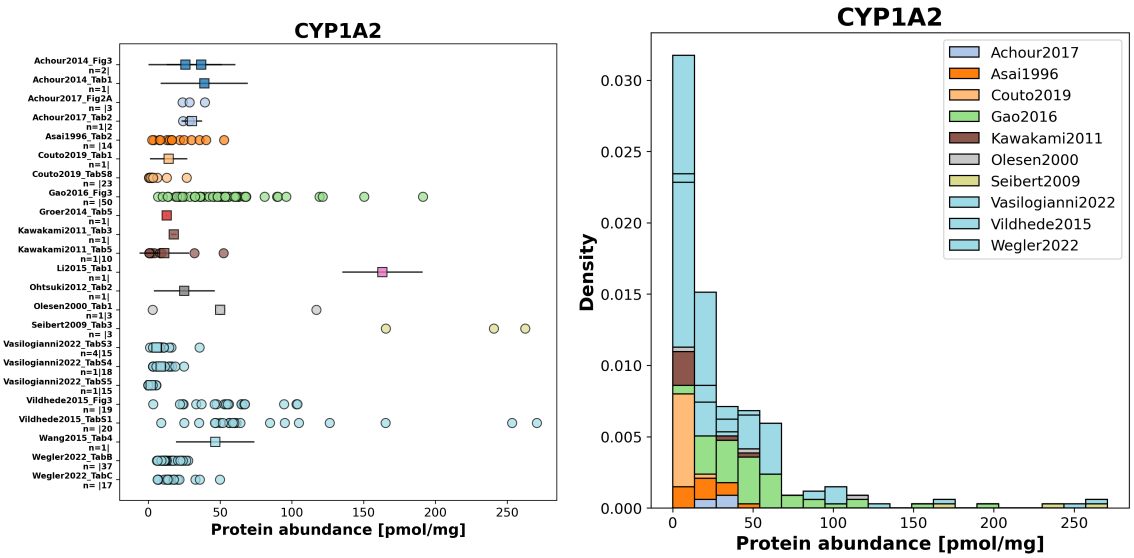


Figure 10: **Protein abundance for CYP1A2.** A) Individual and mean ± SD protein abundance CYP1A2. B) Histogram of individual protein abundance CYP1A2. Data from [2, 1, 6, 10, 15, 17, 25, 27, 29, 30, 33, 39, 40, 41, 42].

3.1.2 Stratification analysis (CYP)

In a next step, we used the established database to investigate factors affecting the protein levels of CYP enzymes in the human liver (Fig. 11). Specifically, we examined the effect of sex (Fig. 12), age (Fig. 13), smoking (Fig. 14), alcohol (Fig. 15), and body mass index (Fig. 16).

category	protein	variable	method	pvalue	sig	pvalue corrected	sig corrected
cyp	CYP4F2	alcohol	Mann-Whitney	1.25638E-05	***	0.001	**
cyp	CYP4F2	age_group	Mann-Whitney	0.001176136	**	0.042	*
cyp	CYP2E1	age_group	Mann-Whitney	0.001546275	**	0.042	*
cyp	CYP2E1	smoking	Mann-Whitney	0.001564649	**	0.042	*
cyp	CYP2B6	age_group	Mann-Whitney	0.001731104	**	0.042	*
cyp	CYP4F2	smoking	Mann-Whitney	0.007030402	**	0.090	
cyp	CYP2C18	age_group	Mann-Whitney	0.007747098	**	0.093	
cyp	CYP3A4	alcohol	Mann-Whitney	0.010308522	*	0.110	
cyp	CYP2C19	age_group	Mann-Whitney	0.011229951	*	0.110	
cyp	CYP2C8	age_group	Mann-Whitney	0.017222338	*	0.141	
cyp	CYP2J2	age_group	Mann-Whitney	0.01851393	*	0.143	
cyp	CYP3A4	smoking	Mann-Whitney	0.02777069	*	0.199	
cyp	CYP3A4	age_group	Mann-Whitney	0.03031725	*	0.209	
cyp	CYP2C19	alcohol	Mann-Whitney	0.035124141	*	0.228	
cyp	CYP3A5	bmi_group	Kruskal-Wallis	0.035480361	*	0.228	
cyp	CYP3A5	age_group	Mann-Whitney	0.041480518	*	0.250	
cyp	CYP2C9	age_group	Mann-Whitney	0.044053732	*	0.250	

Figure 11: **Significance testing of stratification of cytochrome P450 enzymes in human liver.** Significance testing was performed with Mann-Whitney U tests were used in case of two groups (alcohol, age, smoking, sex) and Kruskal-Wallis in case of multiple groups (bmi). Significance levels: *0.05, **0.01, and ***0.001. Rows colored by uncorrected p-values: *** dark green, ** light green * orange. Table contains all entries with uncorrected p-value <0.05.

Significant differences were tested using Mann-Whitney U tests were used in case of two groups (alcohol, age, smoking, sex) and Kruskal-Wallis in case of multiple groups (bmi). P-values were adjusted for multiple testing correction with results presented in Fig. 11.

Age significantly affected CYP4F2, CYP2E1 and CYP2B6, resulting in a decrease in protein amount in the elderly compared to the middle-aged. Smoking had a significant effect on CYP2E1, resulting in a decrease in protein levels. No statistically significant differences in cytochrome P450 levels were found between male and female subjects, or according to BMI or alcohol consumption.

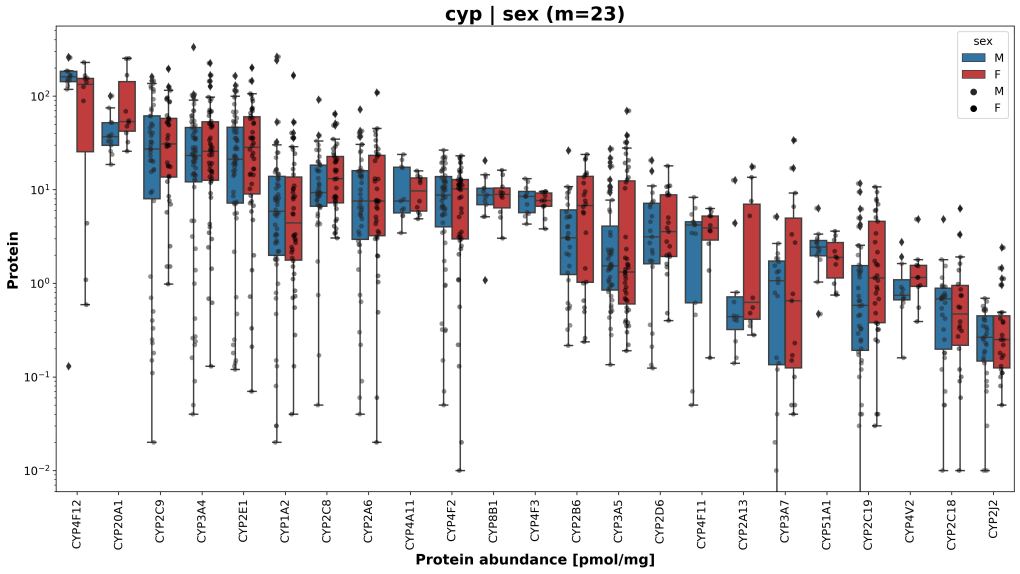


Figure 12: **Stratification of cytochrome P450 enzymes in human liver by sex.** Sufficient data for analysis existed for 23 isoforms.

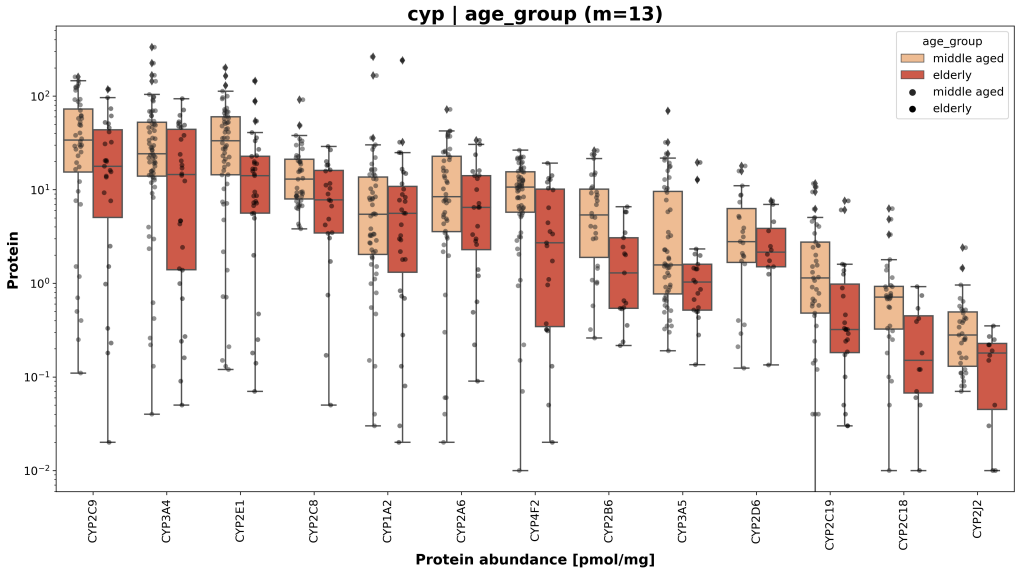


Figure 13: **Stratification of cytochrome P450 enzymes in human liver by age.** Sufficient data for analysis existed for 13 isoforms. Age groups were defined as following: Middle aged (35-65yr), eldery (>65yr)

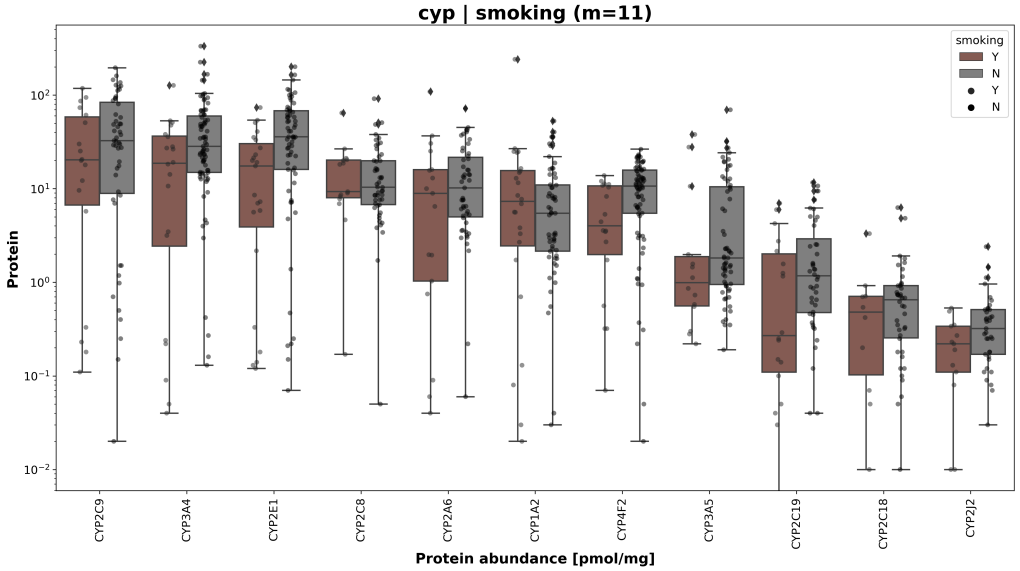


Figure 14: **Stratification of cytochrome P450 enzymes in human liver by smoking status.** Sufficient data for analysis existed for 11 isoforms.

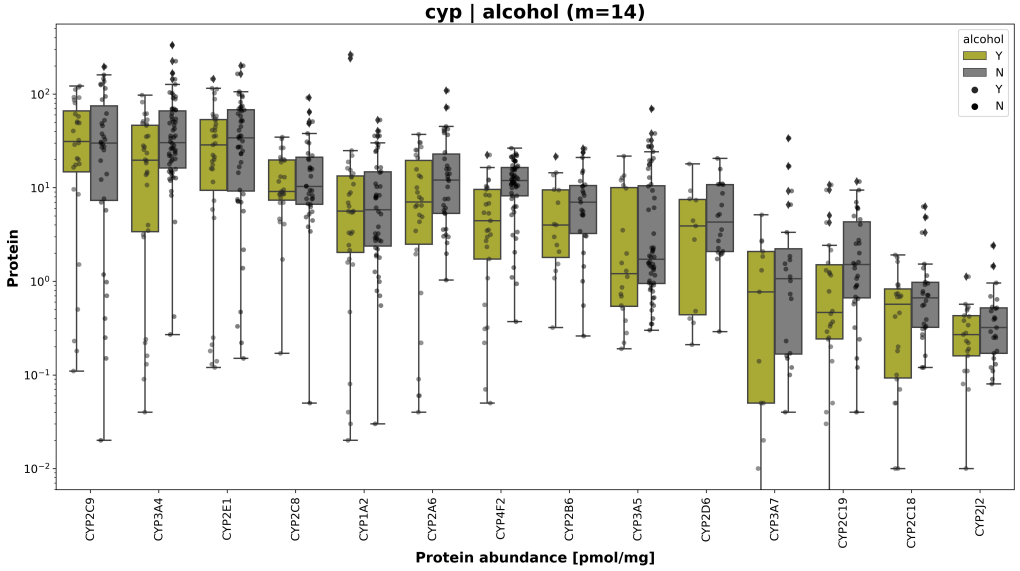


Figure 15: **Stratification of cytochrome P450 enzymes in human liver by alcohol consumption.** Sufficient data for analysis existed for 14 isoforms.

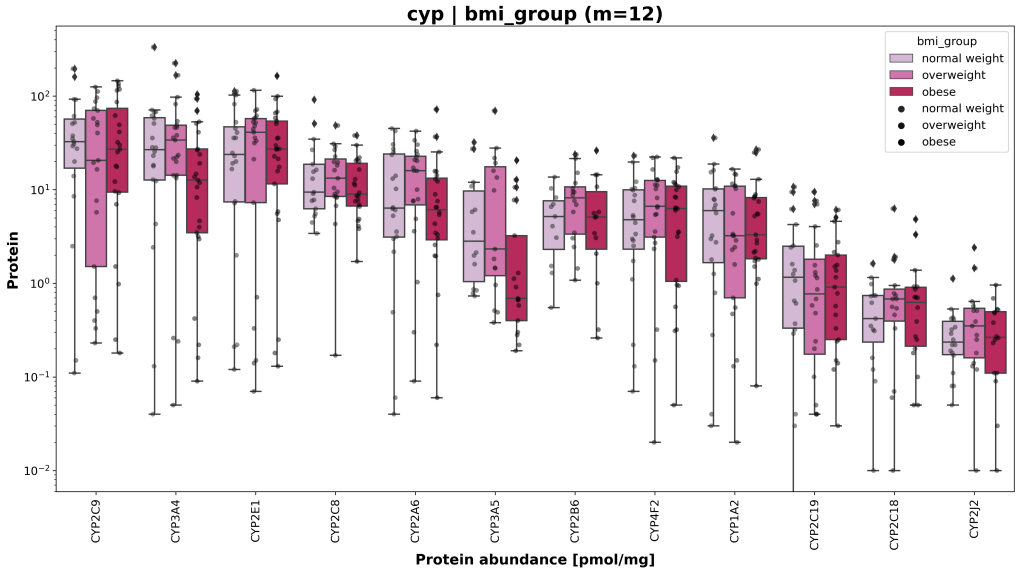


Figure 16: **Stratification of cytochrome P450 enzymes in human liver by body mass index (BMI).** Sufficient data for analysis existed for 12 isoforms. Normal weight (18.5-25 bmi), overweight (25 - 30 bmi), obese (>30 bmi).

3.1.3 Correlation analysis (CYP)

Next, the correlation between different CYP isoforms was analyzed (Fig. 17). The top correlations are shown in a volcano plot (Fig. 18).

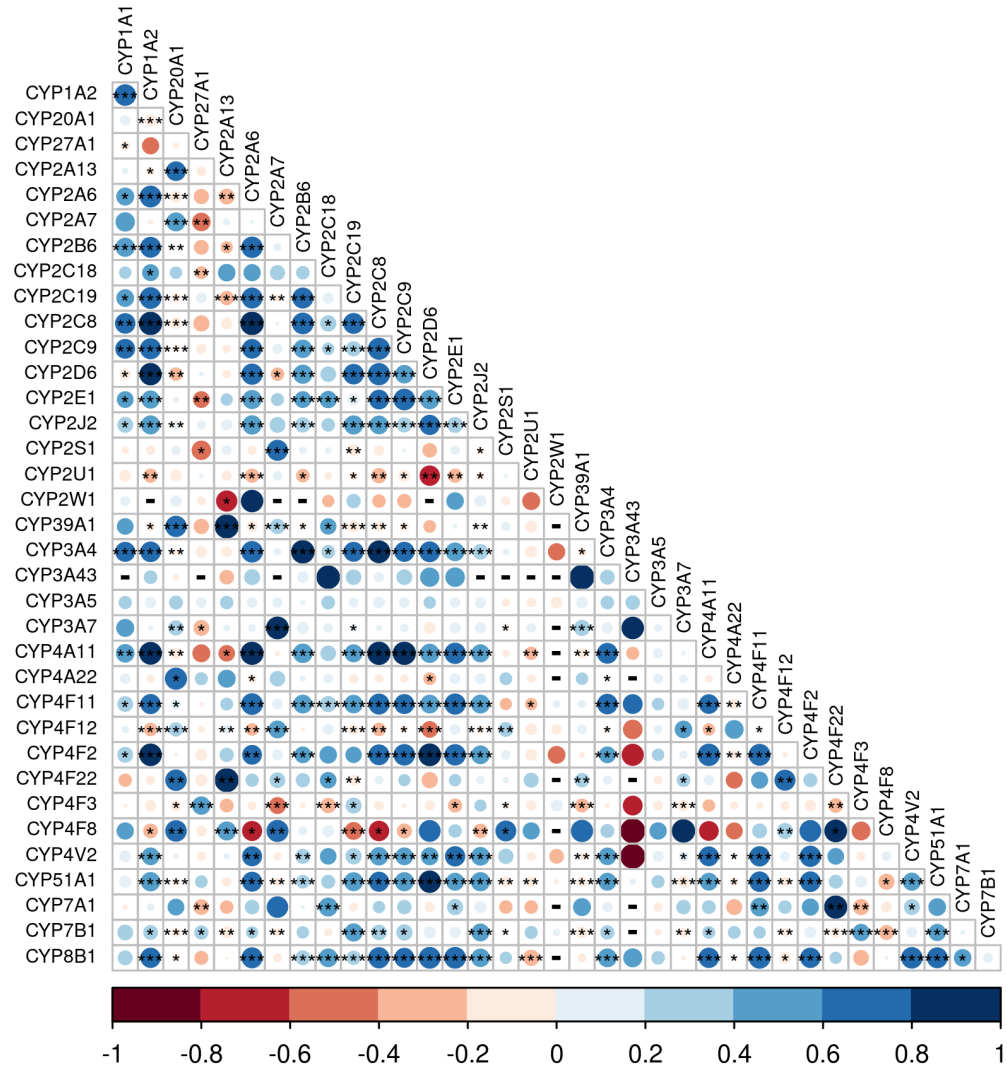


Figure 17: **Correlation matrix of cytochrome P450 isoforms.** Positive correlations are shown in blue and negative correlations are shown in red. The correlation matrix was calculated using Pearson's correlation. Significance levels are *0.05, **0.01, and ***0.001, with p-values adjusted for multiple testing using Benjamini and Hochberg.

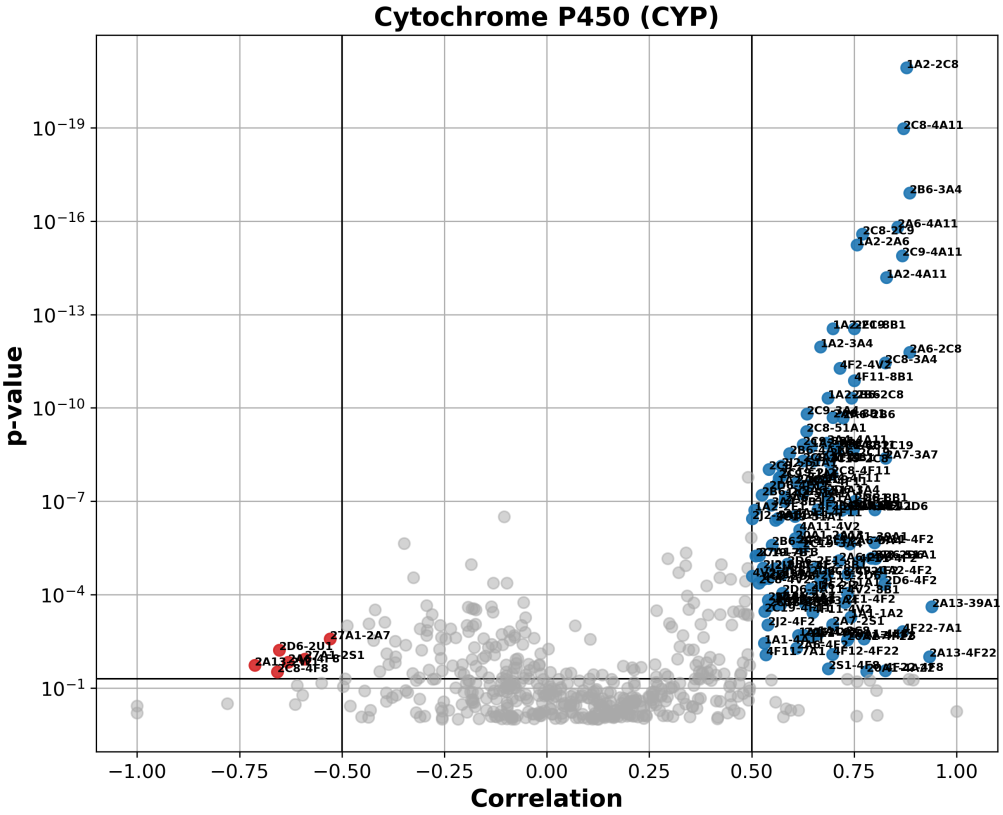


Figure 18: **Volcano plot of correlations between cytochrome P450 isoforms.** Cutoffs for displaying labeled data are p-value <0.05 and absolute correlation coefficient >0.5.

Examples of CYP correlations Examples of the correlation data between different CYPs are given in Fig. 19. One of the strongest positive correlations was found between CYP2C8 and CYP1A2 , and one of the strongest negative correlations was found between CYP2A7 and CYP27A1.

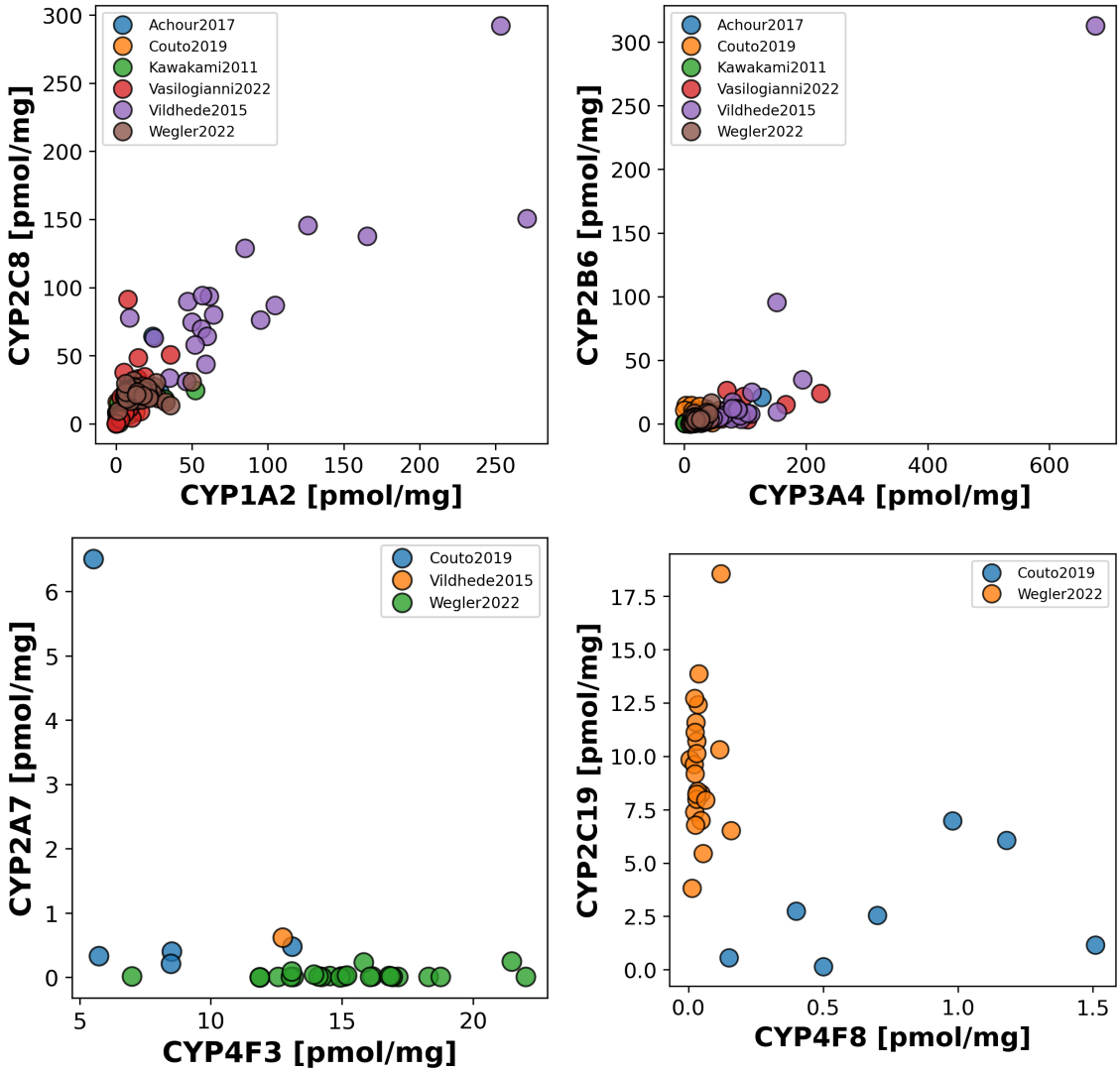


Figure 19: **Protein correlations between cytochrome P450 proteins.** A) Correlation CYP1A2 and CYP2C8. B) Correlation CYP3A4 and CYP2B6. C) Correlation CYP4F3 and CYP2A7. D) Correlation CYP4F8 and CYP2C19. Data from [1, 10, 25, 39, 40, 42].

3.2 UDP-glucuronosyltransferase (UGT)

3.2.1 Protein data (UGT)

An overview of studies containing UGT isoform protein abundances is provided in Fig. 20. In total, 10 studies reported abundance data for 16 UGT isoforms. The resulting protein abundances of the UGT isoforms are shown in Fig. 21 and Fig. 22.

Protein	Achour2017	Couto2019	Groer2014	Li2015	Ohtsuki2012	Takahashi2021	Vasilogianni2022	Vildhede2015	Wang2015	Wegler2022
UGT1A1	|4	1|23	3|	1|	1|	2|	6|32	|40	1|	|91
UGT2B15	|2	1|23	1|	1|	1|	2|	6|42	|26	1|	|61
UGT1A3	|3	1|23	2|	1|	1|	2|	6|46	|40	1|	|79
UGT2B7	|4	1|23	2|	1|	1|	2|	6|50	|39	1|	|91
UGT1A6	|3	1|23		1|	1|	2|	6|39	|40	1|	|85
UGT2B4	|4	1|23		1|		2|	4|20	|40	1|	|63
UGT1A9	|3	1|23		1|	1|	2|	6|47	|20	1|	
UGT2B10	|4	1|23		1|		2|		|40	1|	|60
UGT1A4	|4	1|23		1|		2|		|39	1|	|76
UGT2B17		1|16				2|		|39		|91
UGT3A1		|3						|19		|59
UGT1A8		|1						|19		|3
UGT2A3		|23						|19		|91
UGT2B11		|2					1|14			
UGT2A1		|5								|91
UGT1A7										|55
UGT1A10										|37

Figure 20: **Studies for UGT analysis.** For all studies reporting UGT data, the number of group and individual data points are listed for the different proteins (groups—individuals). UGT data were manually curated from [1, 10, 17, 27, 28, 29, 36, 39, 40, 41, 42].

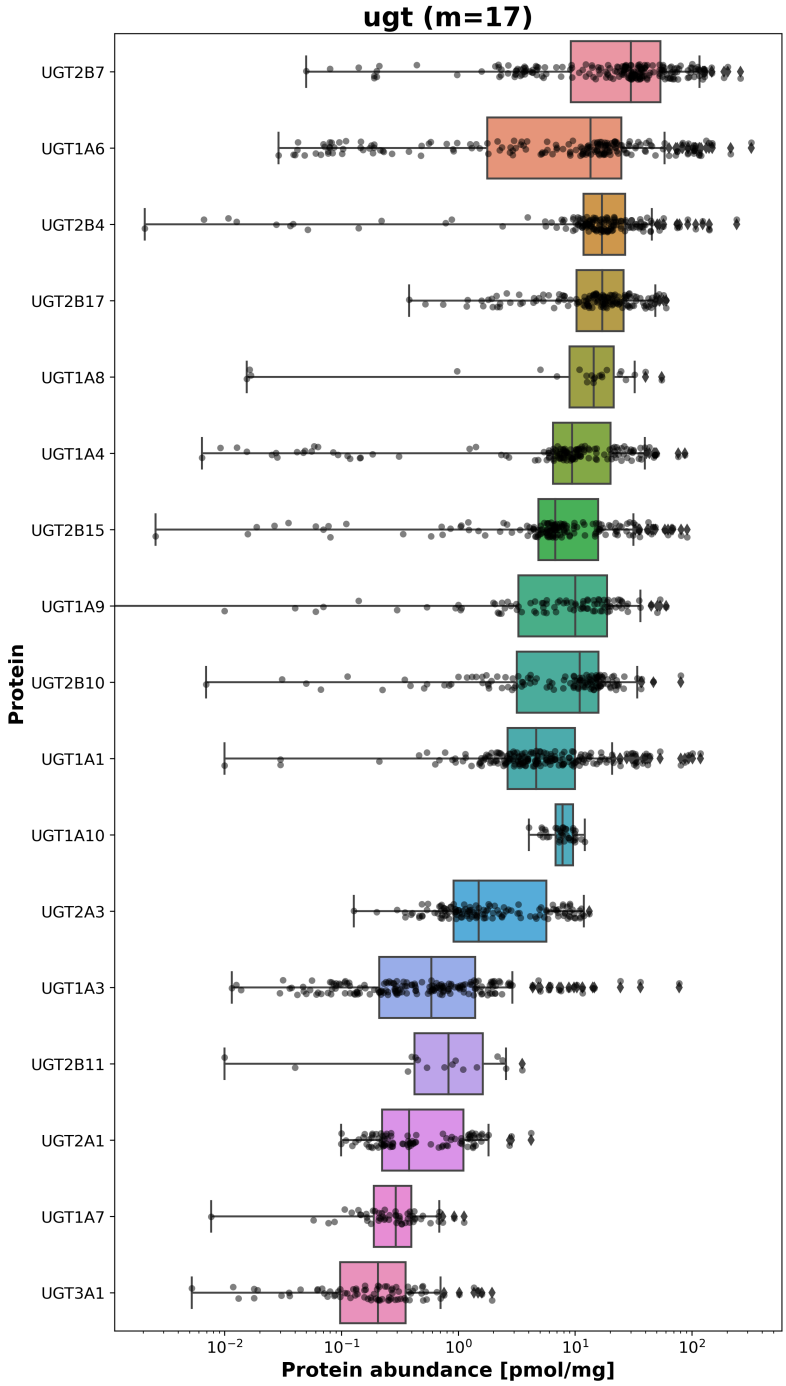


Figure 21: **Protein abundance of UGT isoforms.** The box plot and individual scatter plots are shown for each isoform. Each point corresponds to a single measurement of protein amount in a human liver sample from a single subject.

protein	mean	mode	cv	wcv	min	max	n	unit	Q	I^2	df	heterogeneity
UGT2B7	39.20	0.19	1.03	0.75	0.05	260.91	207	pmol/mg	2.39	0.00	4	none
UGT1A6	26.68	2.93	1.58	0.81	0.03	323.45	190	pmol/mg	11.92	66.43	4	medium
UGT2B4	24.69	10.59	1.19	0.76	0.00	242.81	150	pmol/mg	5.17	22.61	4	none
UGT2B17	19.24	0.38	0.71	0.69	0.38	60.40	146	pmol/mg	1.13	0.00	2	none
UGT1A8	16.49	0.02	0.79	0.49	0.02	55.26	23	pmol/mg		0.00	2	none
UGT1A4	14.13	24.92	1.03	0.71	0.01	86.95	142	pmol/mg	6.73	55.42	3	medium
UGT2B15	12.96	0.00	1.22	0.59	0.00	91.26	154	pmol/mg	88.27	95.47	4	high
UGT1A9	12.79	5.68	0.95	0.85	0.00	60.39	93	pmol/mg	1.34	0.00	3	none
UGT2B10	11.53	0.01	0.92	0.79	0.01	80.27	127	pmol/mg	2.80	0.00	3	none
UGT1A1	10.46	0.03	1.64	0.85	0.01	118.85	190	pmol/mg	2.87	0.00	4	none
UGT1A10	7.89	4.04	0.24	-	4.04	12.11	37	pmol/mg	-	-	0	-
UGT2A3	3.34	0.13	1.02	0.93	0.13	13.26	133	pmol/mg	0.24	0.00	2	none
UGT1A3	2.03	0.24	3.31	1.29	0.01	78.04	191	pmol/mg	9.08	55.97	4	medium
UGT2B11	1.13	0.01	0.88	0.73	0.01	3.54	16	pmol/mg	1.72	41.82	1	low
UGT2A1	0.69	0.10	0.98	0.85	0.10	4.21	96	pmol/mg	0.97	0.00	1	none
UGT1A7	0.33	0.01	0.65	-	0.01	1.12	55	pmol/mg	-	-	0	-
UGT3A1	0.30	0.01	1.16	0.71	0.01	1.94	81	pmol/mg	1.43	0.00	2	none

Figure 22: **Protein abundance and heterogeneity of UGT isoforms.** Weighted means, mode, coefficient of variation (cv), weighted coefficient of variation (wcv), minimum value (min), maximum value (max), number of subjects (n), unit, Cochran's heterogeneity coefficent (Q), Higgen's and Thompson's heterogeneity index (I^2), degree of freedom (df) and heterogeneity analysis of hepatic UGT enzyme abundance data.

Examples of protein levels (UGT) Two examples are provided with protein abundances for UGT2B7 shown in Fig. 23 and UGT1A1 in Fig. 24.

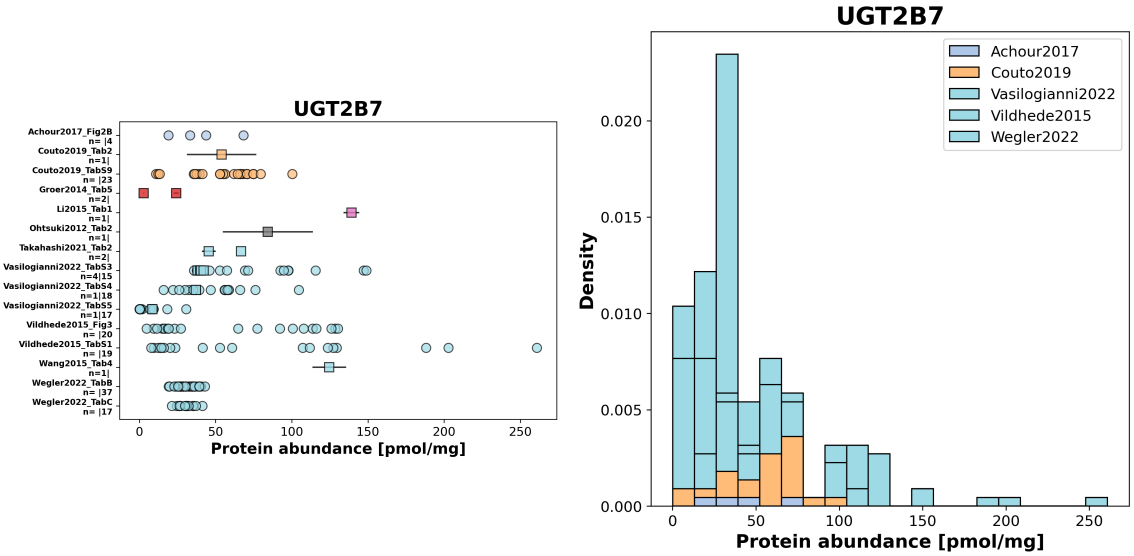


Figure 23: **Protein abundance for UGT2B7.** A) Individual and mean ± SD protein abundance UGT2B7. B) Histogram of individual protein abundance UGT2B7. Data from [1, 10, 17, 27, 29, 36, 39, 40, 41, 42].

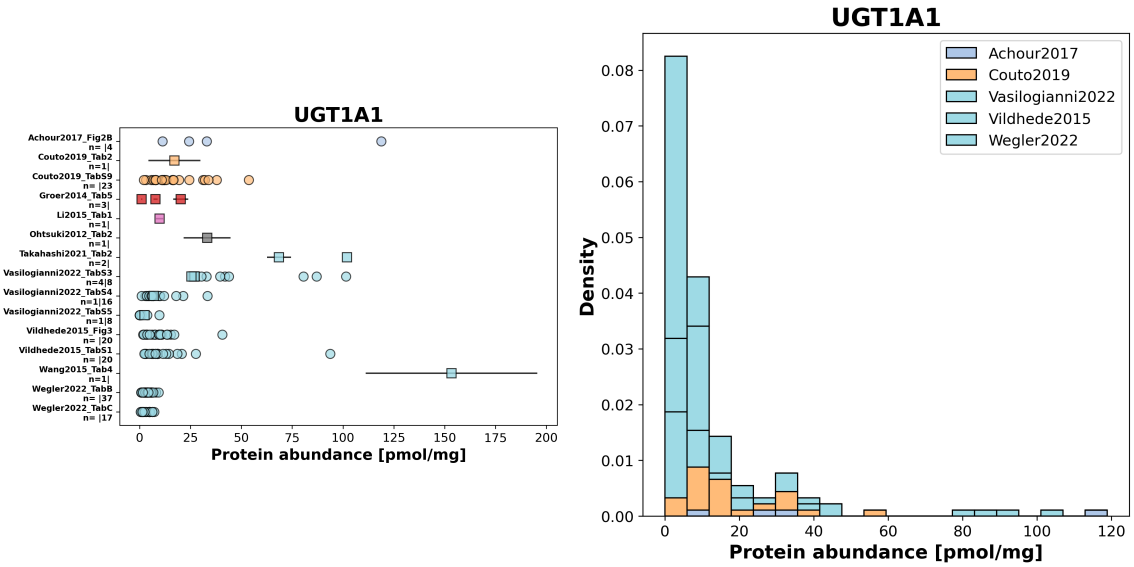


Figure 24: **Protein abundance for UGT1A1.** A) Individual and mean ± SD protein abundance UGT1A1. B) Histogram of individual protein abundance UGT1A1. Data from [1, 10, 17, 27, 29, 36, 39, 40, 41, 42].

3.2.2 Stratification analysis (UGT)

In a next step, we used the established database to investigate factors affecting the protein levels of UGT enzymes in the human liver (Fig. 25). Specifically, we examined the effect of sex (Fig. 26), age (Fig. 27), smoking (Fig. 28), alcohol (Fig. 29), and body mass index (Fig. 30).

category	protein	variable	method	pvalue	sig	pvalue corrected	sig corrected
ugt	UGT1A6	age_group	Mann-Whitney	0.00036681351	***	0.01769875176419	*
ugt	UGT1A3	smoking	Mann-Whitney	0.00303678868	**	0.05861002159496	
ugt	UGT1A3	age_group	Mann-Whitney	0.00463889773	**	0.06886978931836	
ugt	UGT2B7	age_group	Mann-Whitney	0.00915549047	**	0.10394174479708	
ugt	UGT2B7	alcohol	Mann-Whitney	0.0119742145	*	0.11004873330062	
ugt	UGT2B7	smoking	Mann-Whitney	0.02280832911	*	0.16930798148329	
ugt	UGT1A9	smoking	Mann-Whitney	0.04311074562	*	0.25006971285944	

Figure 25: **Significance testing of stratification of UGT enzymes in human liver.** Significance testing was performed with Mann-Whitney U tests were used in case of two groups (alcohol, age, smoking, sex) and Kruskal-Wallis in case of multiple groups (bmi). Significance levels: *0.05, **0.01, and ***0.001. Rows colored by uncorrected p-values: *** dark green, ** light green * orange. Table contains all entries with uncorrected p-value <0.05.

Significant differences were tested using Mann-Whitney U tests were used in case of two groups (alcohol, age, smoking, sex) and Kruskal-Wallis in case of multiple groups (bmi). P-values were adjusted for multiple testing correction with results presented in Fig. 25.

Age had a significant effect on UGT1A6, resulting in a significant decrease in protein amount in elderly compared to middle aged subjects. No statistically significant differences in cytochrome P450 abundance levels were found with sex, smoking, alcohol consumption or body mass index.

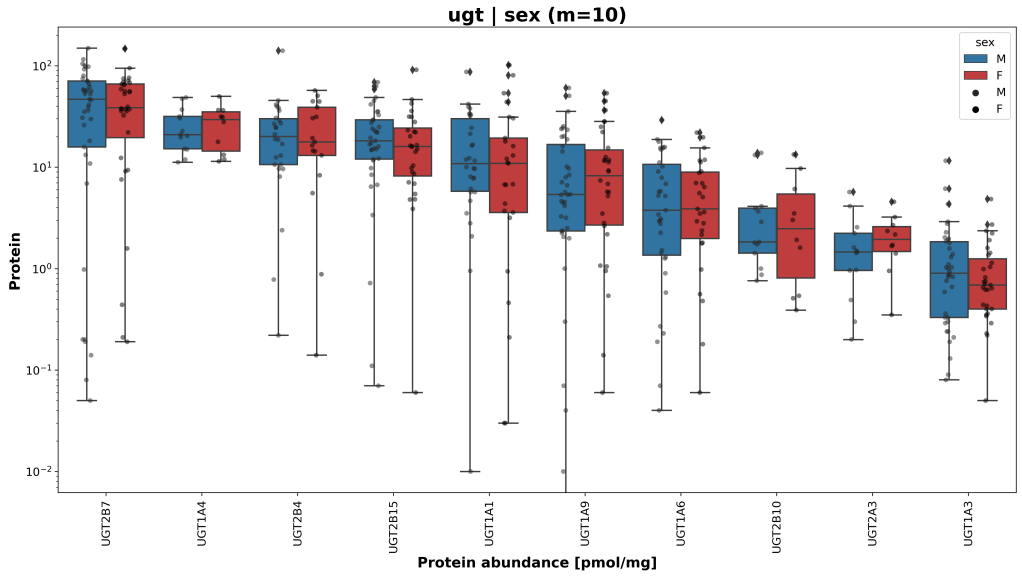


Figure 26: **Stratification of UGT enzymes in human liver by sex.** Sufficient data for analysis existed for 10 isoforms.

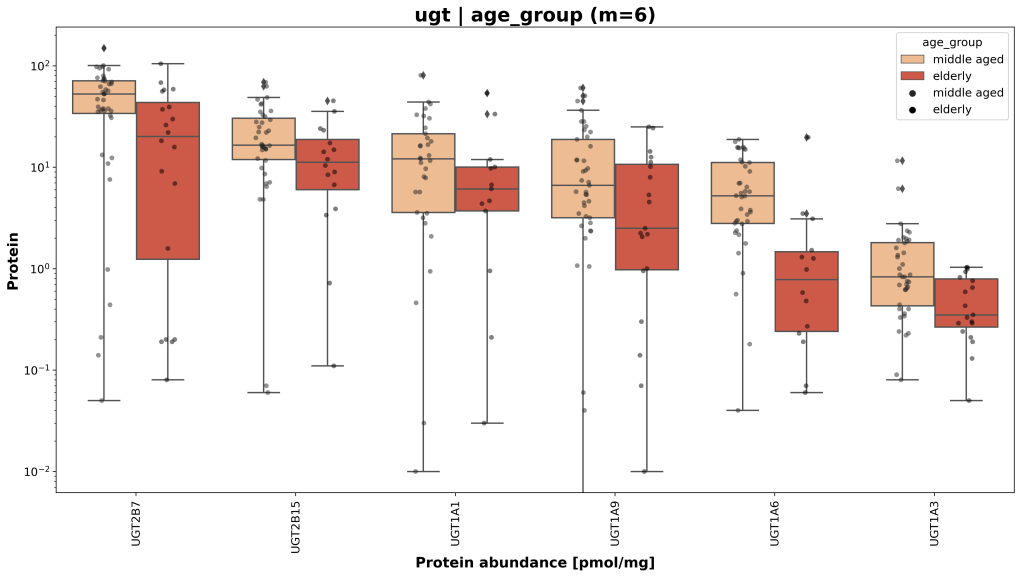


Figure 27: **Stratification of UGT enzymes in human liver by age.** Sufficient data for analysis existed for 6 isoforms. Age groups were defined as following: Middle aged (35-65yr), eldery (>65yr)

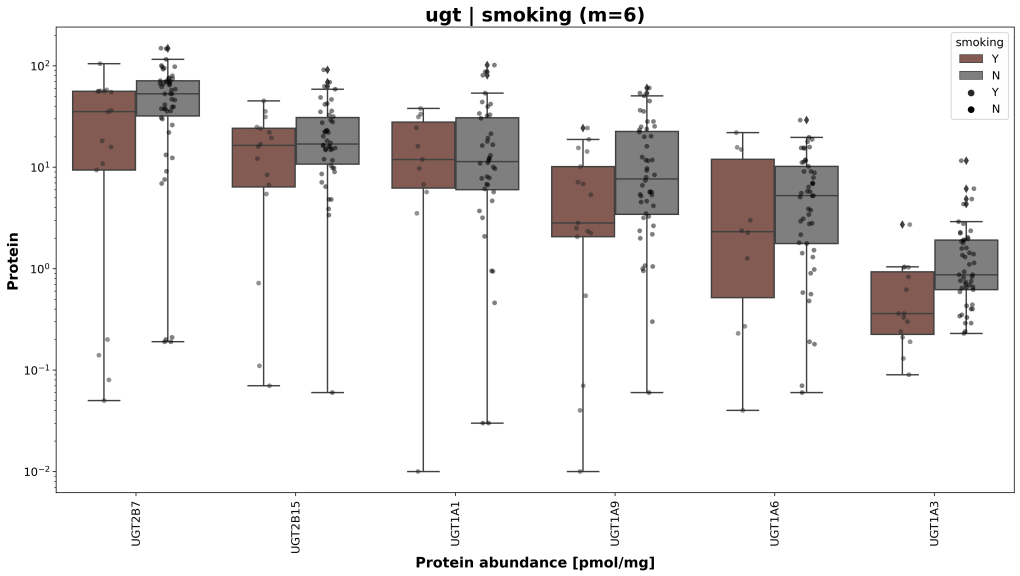


Figure 28: **Stratification of UGT enzymes in human liver by smoking status.** Sufficient data for analysis existed for 6 isoforms.

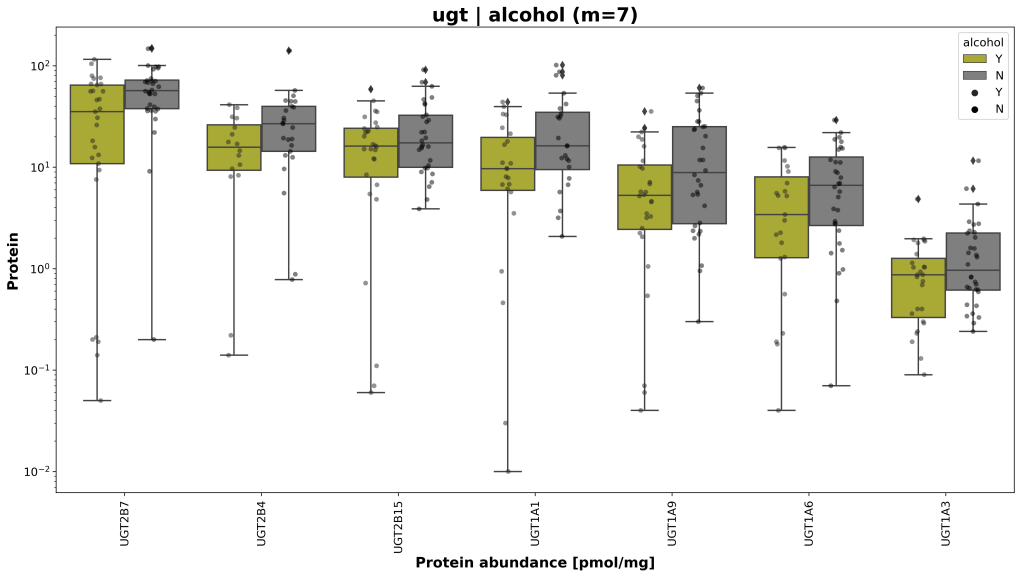


Figure 29: **Stratification of UGT enzymes in human liver by alcohol consumption.** Sufficient data for analysis existed for 7 isoforms.

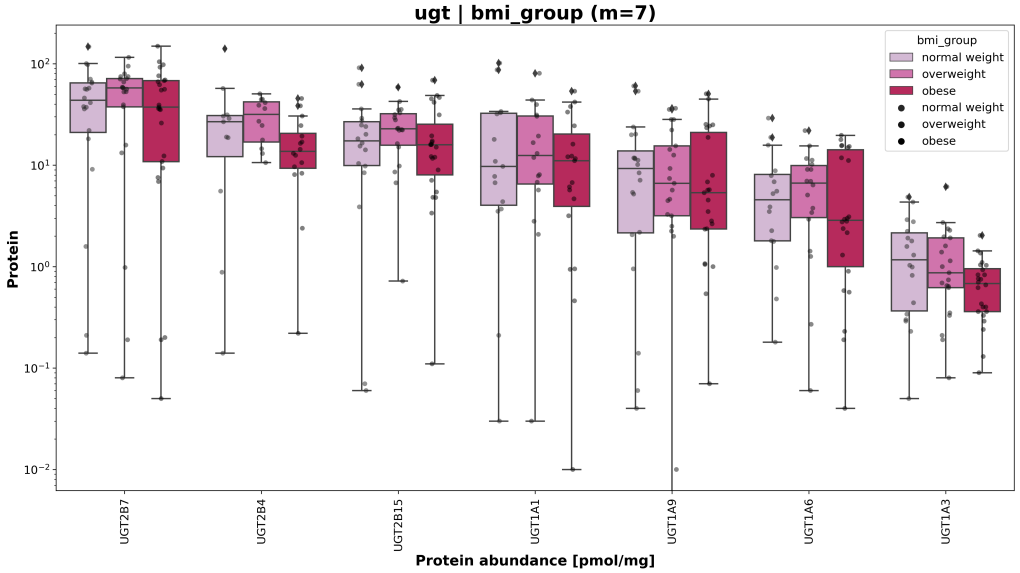


Figure 30: **Stratification of UGT enzymes in human liver by body mass index (BMI).** Sufficient data for analysis existed for 7 isoforms. Normal weight (18.5-25 bmi), overweight (25 - 30 bmi), obese (>30 bmi).

3.2.3 Correlation analysis (UGT)

Next, the correlation between different UGT isoforms was analyzed (Fig. 31). The top correlations are shown in a volcano plot (Fig. 32).

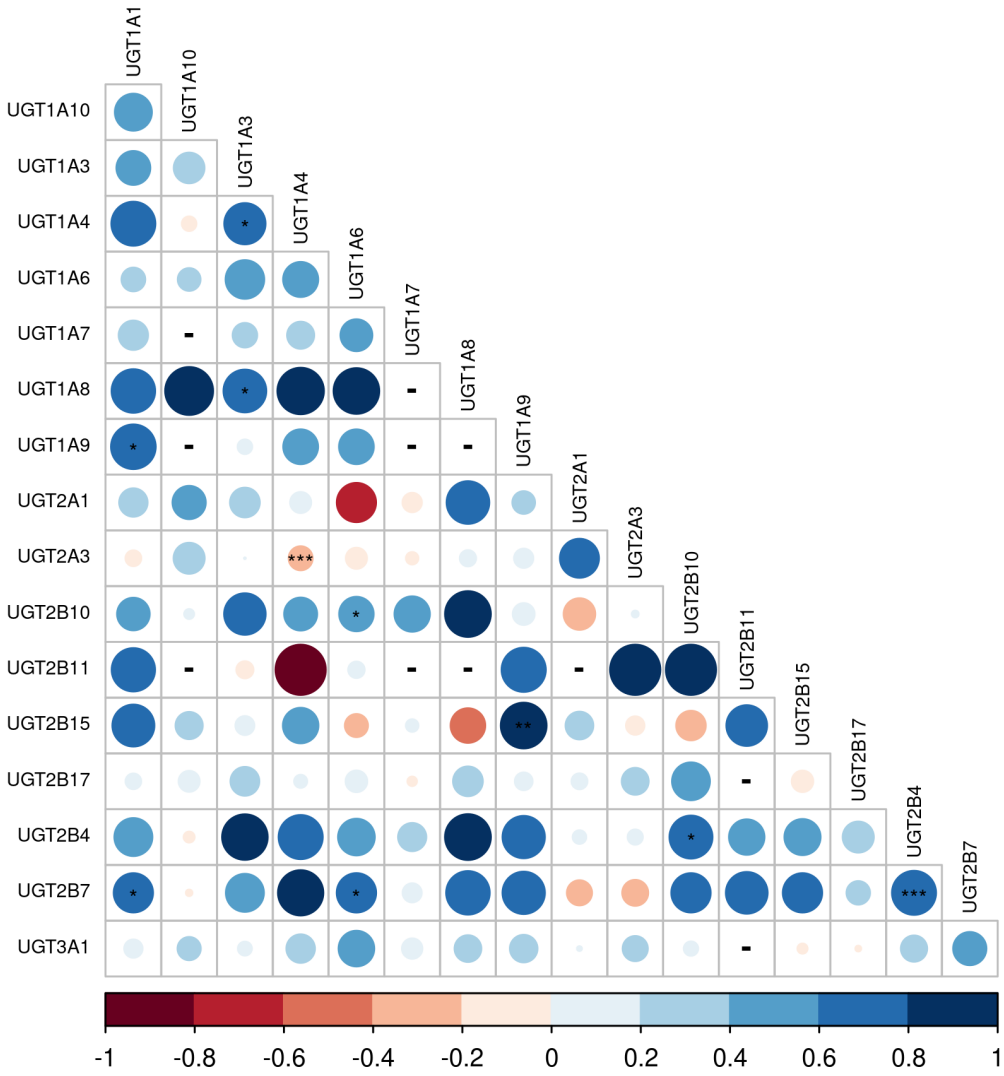


Figure 31: **Correlation matrix of UGT isoforms.** Positive correlations are shown in blue and negative correlations are shown in red. The correlation matrix was calculated using Pearson's correlation. Significance levels are *0.05, **0.01, and ***0.001, with p-values adjusted for multiple testing using Benjamini and Hochberg.

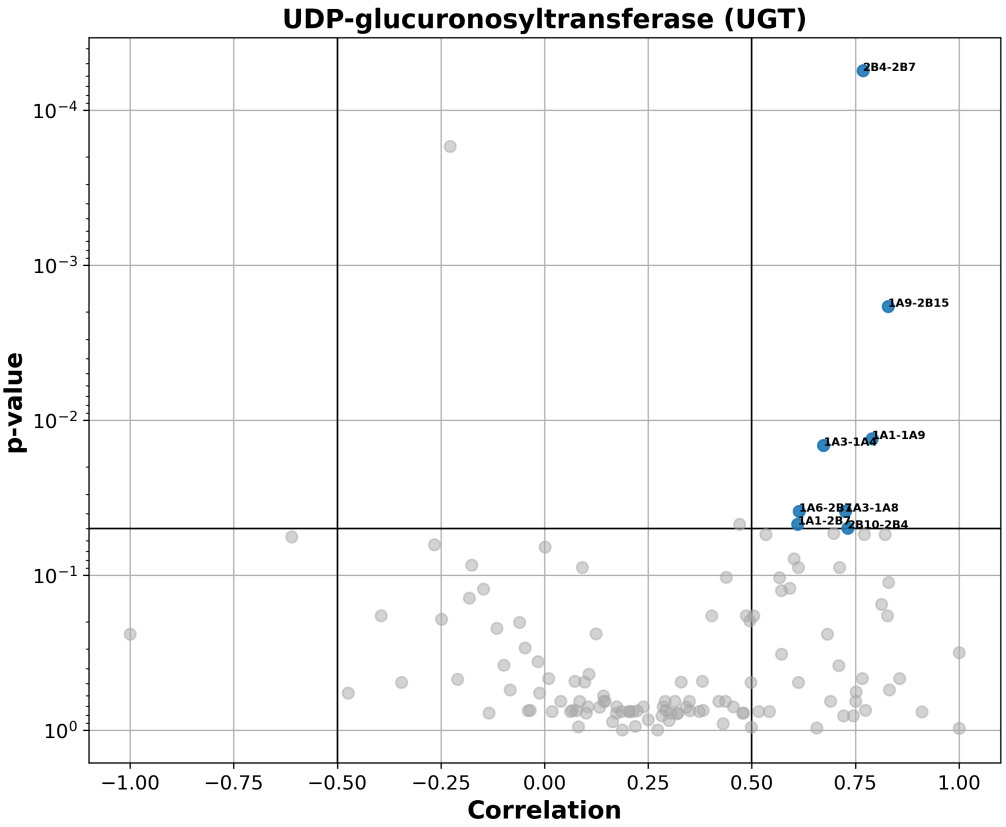


Figure 32: **Volcano plot of correlations between UGT isoforms.** Cutoffs for displaying labeled data are p-value <0.05 and absolute correlation coefficient >0.5.

Examples of UGT correlations Examples of the correlation data between different UGTs are given in Fig. 33.

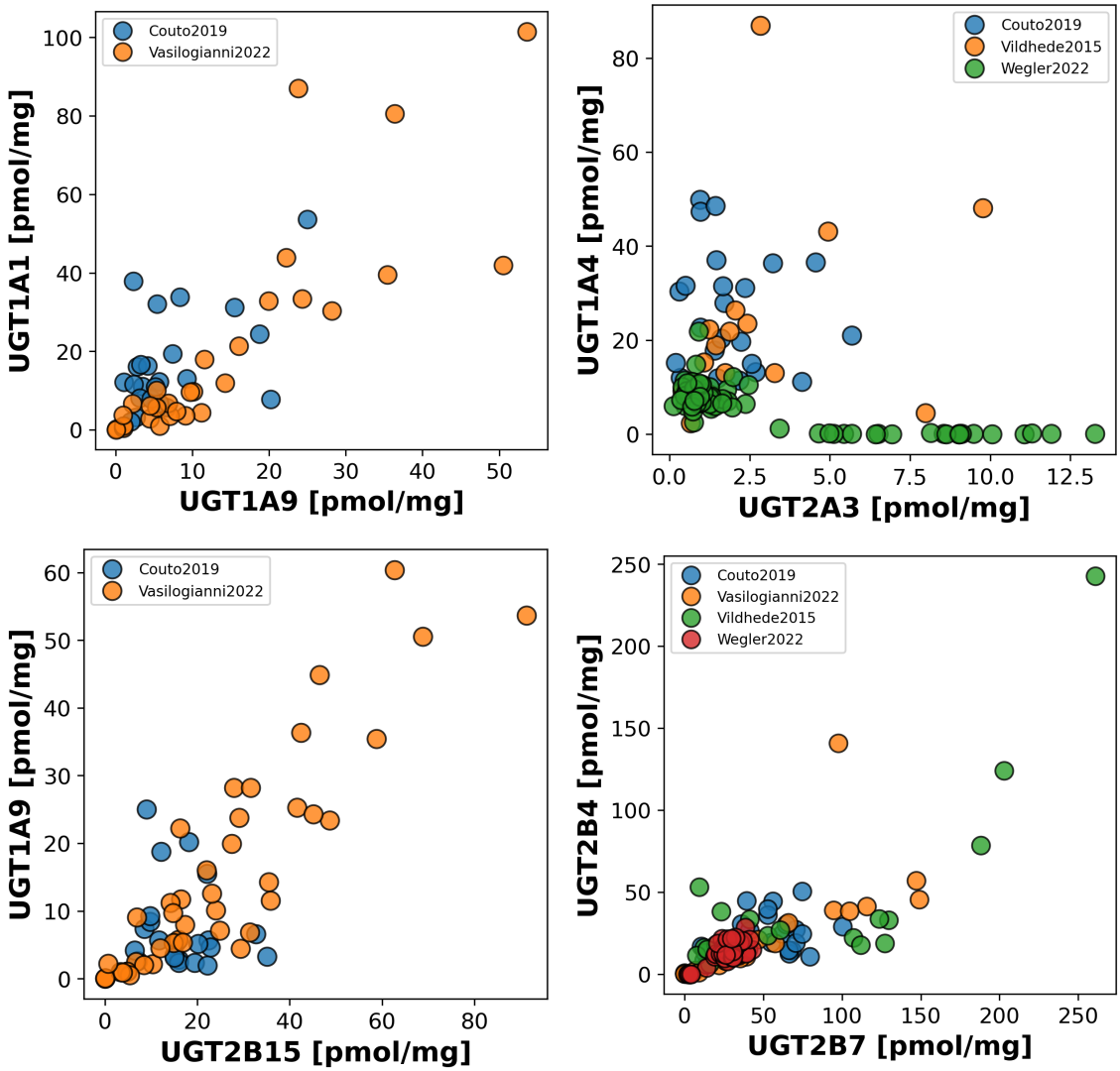
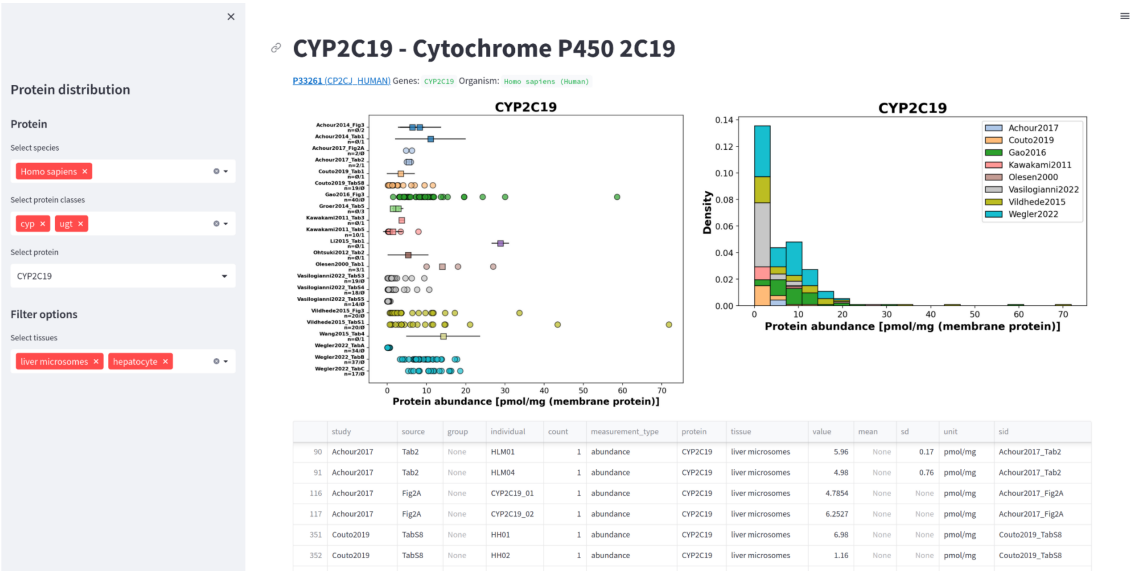


Figure 33: **Protein correlations between cytochrome P450 proteins.** A) Correlation UGT1A9 and UGT1A1. B) Correlation UGT2A3 and UGT1A4. C) Correlation UGT2B15 and UGT1A9. D) Correlation UGT2B7 and UGT2B4. Data from [10, 39, 40, 42].

3.3 Data base and web tool

The data is freely available via a web application from https://protein-distribution.streamlit.app/. The web application allows to browse the data for the individual CYP and UGT isoforms and can be used for interactive data curation. Abundance data can be downloaded as CSV.



	study	source	group	individual	count	measurement_type	protein	tissue	value	mean	sd	unit	sid
90	Achour2017	Tab2	None	HLM01	1	abundance	CYP2C19	liver microsomes	5.96	None	0.17	pmol/mg	Achour2017_Tab2
91	Achour2017	Tab2	None	HLM04	1	abundance	CYP2C19	liver microsomes	4.98	None	0.76	pmol/mg	Achour2017_Tab2
116	Achour2017	Fig2A	None	CYP2C19_01	1	abundance	CYP2C19	liver microsomes	4.7854	None	None	pmol/mg	Achour2017_Fig2A
117	Achour2017	Fig2A	None	CYP2C19_02	1	abundance	CYP2C19	liver microsomes	6.2527	None	None	pmol/mg	Achour2017_Fig2A
351	Couto2019	TabS8	None	HH01	1	abundance	CYP2C19	liver microsomes	6.98	None	None	pmol/mg	Couto2019_TabS8
352	Couto2019	TabS8	None	HH02	1	abundance	CYP2C19	liver microsomes	1.16	None	None	pmol/mg	Couto2019_TabS8

Figure 34: **Overview protein web application.** Protein isoforms can be selected and the corresponding protein abundance data is displayed. Data can be downloaded as CSV.

3.4 Summary

In summary, this work has established the first free and open database of CYP and UGT protein levels in the human liver. Analysis of the data revealed that: (i) CYP and UGT isoforms show large inter-individual variability in protein levels. (ii) Different CYP and UGT isoforms have different protein levels. (iii) Factors such as age, smoking and alcohol consumption may affect the protein levels of CYP and UGT enzymes. (iv) There are significant correlations in protein levels between CYP isoforms and between UGT isoforms.

This work has established an important resource for the study of variability in drug detoxification due to variability in protein levels in the human liver.

4 Discussion

4.1 Data

This work established the first free and open database of CYP and UGT protein levels in human liver. This in itself is a unique resource for studying cytochrome and uridine 5'-diphosphoglucuronosyltransferase metabolism and its protein variability. To create the datasets curation and evaluation of eighteen selected CYP studies and ten selected UGT studies were required. The publication date of the curated studies range from 1954 to 2023. The database allows us to study important questions, such as possible factors influencing protein levels or possible correlation of drug-metabolizing enzymes in human liver. The established database has several limitations. The majority of the data was collected from Caucasian individuals, with other ethnicities underrepresented. Therefore, the analysis and results are for Caucasian subjects only. Stratification analysis based on ethnicity was not possible. The data of the few subjects categorized in the younger age group (¡35) could not be included for the analysis. Importantly, the data were used as reported in the various studies. It is well known that datasets from different research groups and publications often show systematic bias. For example, in a kinetic model of the pancreas, this could be observed for the metabolic data [11] , and methods to remove the systematic bias significantly improved the quality of the integrated dataset. Similar things are observed for example in omics analysis [16] and methods are applied to remove the differences between the datasets [3, 38]]. In our study, no correction for confounding study factors and systematic bias of the data was applied. The results of the subsequent analyses all depend on the quality of the underlying data, which may be a limiting factor. Some of the large inter-individual variability observed in the data is likely due to between-study variability. We expect that removing this bias will reduce this variability and provide more meaningful results when analyzing factors that affect protein levels.

For example, smoking is known to induce CYP1A2 [18] and the trend can be seen in the data, but it was not significant. Removing the technical variability due to between-study variability may allow better discrimination of biological variability and factors affecting protein levels, such as smoking on CYP1A2.

4.2 Analysis

The observed large inter-individual variability in CYP and UGT isoforms in the human liver will have significant implications for variability in drug metabolism. For example, recent work has demonstrated the large effect of variability in CYP3A4 and CYP2D6 on dextromethorphan metabolism and metabolic phenotyping [20]. While the effect of genetic polymorphisms on drug metabolism and pharmacokinetics is often discussed, e.g. for CYP2C19 (omeprazole) or CYP2D6 (dextromethorphan), the large effect of variability in protein levels is often ignored. Women have higher protein levels of CYP3A4 than men [31], according to a recent study, which shows up moderately in our data. An important finding of the analysis is that multiple factors can affect the protein levels of CYPs and UGTs in the human liver. Age had a significant effect on CYP4F2, CYP2E1, CYP2B6 and UGT1A6, resulting in significantly lower protein levels in elderly compared to middle-aged individuals. This may have important implications for drug dose adjustment in the elderly for drugs that are metabolized by these enzymes. Younger aged subjects are not included in the data.

Smoking had a significant effect on CYP2E1, resulting in a decrease in protein levels. As a result, smoking will affect the metabolism of drugs and substances affected by CYP2E1, such as chlorzoxazone or ethanol.

TODO: BMI, Sex

5 Outlook

This study is an important first step in unraveling the complexity of protein variability in drug metabolism, with our analysis providing a critical understanding of the role of cytochrome P450 and UGT enzymes in the human liver.

Future work will focus on eliminating systematic biases in our study methods. The establishment of a reproducible workflow in this research paves the way for seamless integration with the analyses performed.

Although our current dataset is a significant contribution, it represents only a fraction of the proteins involved in drug metabolism in the human liver. The dataset is being actively expanded to include a wider range of enzymes, in particular ABC and SLC transporters. In addition, we aim to extend our research to other critical organs such as the kidney and intestine, and to other species such as rats and mice.

This data set is a critical resource for computational modeling. Upcoming research will utilize this data to explore the variation in drug metabolism due to variability in CYP and UGT enzymes. This investigation will use physiologically based pharmacokinetic (PBPK) models to improve our understanding of pharmacodynamics in different biological systems.

6 Contributions

Afruja Hossain (AH) performed the data curation and data analysis consisting of protein analyis, stratification analysis and correlation analysis. AH wrote the thesis and created the figures and tables.

Matthias König (MK) supervised the project, implemented the web application, and performed the statistical analysis. MK functioned as a second data curator ensuring consistency of the curated data.

Acknowledgements

This work was supported by the Federal Ministry of Education and Research (BMBF, Germany) within ATLAS by grant number 031L0304B and by the German Research Foundation (DFG) within the Research Unit Program FOR 5151 "QuaLiPerF (Quantifying Liver Perfusion-Function Relationship in Complex Resection - A Systems Medicine Approach)" by grant number 436883643 and by grant number 465194077 (Priority Programme SPP 2311, Subproject SimLivA).

References

[1] Brahim Achour, Hajar Al Feteisi, Francesco Lanucara, Amin Rostami-Hodjegan, and Jill Barber. "Global Proteomic Analysis of Human Liver Microsomes: Rapid Characterization and Quantification of Hepatic Drug-Metabolizing Enzymes". In: *Drug Metabolism and Disposition* 45.6 (June 2017), pp. 666–675. DOI: 10.1124/dmd.116.074732. URL: http://dmd.aspetjournals.org/lookup/doi/10.1124/dmd.116.074732 (visited on 10/31/2022).

[2] Brahim Achour, Jill Barber, and Amin Rostami-Hodjegan. "Expression of Hepatic Drug-Metabolizing Cytochrome P450 Enzymes and Their Intercorrelations: A Meta-Analysis". In: *Drug Metabolism and Disposition* 42.8 (Aug. 2014), pp. 1349–1356. DOI: 10.1124/dmd.114.058834. URL: http://dmd.aspetjournals.org/lookup/doi/10.1124/dmd.114.058834 (visited on 06/10/2022).

[3] Michael F. Adamer, Sarah C. Brüningk, Alejandro Tejada-Arranz, Fabienne Estermann, Marek Basler, and Karsten Borgwardt. "reComBat: Batch-Effect Removal in Large-Scale Multi-Source Gene-Expression Data Integration". In: *Bioinformatics Advances* 2.1 (2022), vbac071. DOI: 10.1093/bioadv/vbac071. pmid: 36699372.

[4] Omar Abdulhameed Almazroo, Mohammad Kowser Miah, and Raman Venkataramanan. "Drug Metabolism in the Liver". In: *Clinics in Liver Disease* 21.1 (Feb. 2017), pp. 1–20. DOI: 10.1016/j.cld.2016.08.001. pmid: 27842765.

[5] Peter Armitage, Geoffrey Berry, and John Nigel Scott Matthews. *Statistical Methods in Medical Research*. John Wiley & Sons, 2008.

[6] H. Asai, S. Imaoka, T. Kuroki, T. Monna, and Y. Funae. "Microsomal Ethanol Oxidizing System Activity by Human Hepatic Cytochrome P450s". In: *The Journal of Pharmacology and Experimental Therapeutics* 277.2 (May 1996), pp. 1004–1009. pmid: 8627510.

[7] E. Boyland and L. F. Chasseaud. "The Role of Glutathione and Glutathione S-transferases in Mercapturic Acid Biosynthesis". In: *Advances in Enzymology and Related Areas of Molecular Biology* 32 (1969), pp. 173–219. DOI: 10.1002/9780470122778.ch5. pmid: 4892500.

[8] William G. Cochran. "The Combination of Estimates from Different Experiments". In: *Biometrics* 10.1 (Mar. 1954), p. 101. DOI: 10.2307/3001666. JSTOR: 3001666. URL: https://www.jstor.org/stable/3001666?origin=crossref (visited on 12/21/2023).

[9] J. N. Commandeur, G. J. Stijntjes, and N. P. Vermeulen. "Enzymes and Transport Systems Involved in the Formation and Disposition of Glutathione S-conjugates. Role in Bioactivation and Detoxication Mechanisms of Xenobiotics". In: *Pharmacological Reviews* 47.2 (June 1995), pp. 271–330. pmid: 7568330.

[10] Narciso Couto, Zubida M. Al-Majdoub, Brahim Achour, Phillip C. Wright, Amin Rostami-Hodjegan, and Jill Barber. "Quantification of Proteins Involved in Drug Metabolism and Disposition in the Human Liver Using Label-Free Global Proteomics". In: *Molecular Pharmaceutics* 16.2 (Feb. 4, 2019), pp. 632–647. DOI: 10.1021/acs.molpharmaceut.8b00941. URL: https://pubs.acs.org/doi/10.1021/acs.molpharmaceut.8b00941 (visited on 08/17/2022).

[11] M. Deepa Maheshvare, Soumyendu Raha, Matthias König, and Debnath Pal. "A Pathway Model of Glucose-Stimulated Insulin Secretion in the Pancreatic Beta-Cell". In: *Frontiers in Endocrinology* 14 (2023), p. 1185656. DOI: 10.3389/fendo.2023.1185656. pmid: 37600713.

[12] Hartmut Derendorf, Stephan Schmidt, Malcolm Rowland, and Thomas N. Tozer. *Rowland and Tozer's Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications*. Fifth edition. Philadelphia: Wolters Kluwer, 2020. 939 pp.

[13] S. N. de Wildt, G. L. Kearns, J. S. Leeder, and J. N. van den Anker. "Cytochrome P450 3A: Ontogeny and Drug Disposition". In: *Clinical Pharmacokinetics* 37.6 (Dec. 1999), pp. 485–505. DOI: 10.2165/00003088-199937060-00004. pmid: 10628899.

[14] W. E. Evans and M. V. Relling. "Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics". In: *Science (New York, N.Y.)* 286.5439 (Oct. 15, 1999), pp. 487–491. DOI: 10.1126/science.286.5439.487. pmid: 10521338.

[15] Na Gao, Xin Tian, Yan Fang, Jun Zhou, Haifeng Zhang, Qiang Wen, Linjing Jia, Jie Gao, Bao Sun, Jingyao Wei, Yunfei Zhang, Mingzhu Cui, and Hailing Qiao. "Gene Polymorphisms and Contents of Cytochrome P450s Have Only Limited Effects on Metabolic Activities in Human Liver Microsomes". In: *European Journal of Pharmaceutical Sciences* 92 (Sept. 2016), pp. 86–97. DOI: 10.1016/j.ejps.2016.06.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0928098716302299 (visited on 06/14/2022).

[16] Wilson Wen Bin Goh, Wei Wang, and Limsoon Wong. "Why Batch Effects Matter in Omics Data, and How to Avoid Them". In: *Trends in Biotechnology* 35.6 (June 2017), pp. 498–507. DOI: 10.1016/j.tibtech.2017.02.012. pmid: 28351613.

[17] C. Gröer, D. Busch, M. Patrzyk, K. Beyer, A. Busemann, C. D. Heidecke, M. Drozdzik, W. Siegmund, and S. Oswald. "Absolute Protein Quantification of Clinically Relevant Cytochrome P450 Enzymes and UDP-glucuronosyltransferases by Mass Spectrometry-Based Targeted Proteomics". In: *Journal of Pharmaceutical and Biomedical Analysis* 100 (Nov. 2014), pp. 393–401. DOI: 10.1016/j.jpba.2014.08.016. pmid: 25218440.

[18] Jan Grzegorzewski, Florian Bartsch, Adrian Köller, and Matthias König. "Pharmacokinetics of Caffeine: A Systematic Analysis of Reported Data for Application in Metabolic Phenotyping and Liver Function Testing". In: *Frontiers in Pharmacology* 12 (2021), p. 752826. DOI: 10.3389/fphar.2021.752826. pmid: 35280254.

[19] Jan Grzegorzewski, Janosch Brandhorst, Kathleen Green, Dimitra Eleftheriadou, Yannick Duport, Florian Barthorscht, Adrian Köller, Danny Yu Jia Ke, Sara De Angelis, and Matthias König. "PK-DB: Pharmacokinetics Database for Individualized and Stratified Computational Modeling". In: *Nucleic Acids Research* 49.D1 (Jan. 8, 2021), pp. D1358–D1364. DOI: 10.1093/nar/gkaa990. pmid: 33151297.

[20] Jan Grzegorzewski, Janosch Brandhorst, and Matthias König. "Physiologically Based Pharmacokinetic (PBPK) Modeling of the Role of CYP2D6 Polymorphism for Metabolic Phenotyping with Dextromethorphan". In: *Frontiers in Pharmacology* 13 (2022), p. 1029073. DOI: 10.3389/fphar.2022.1029073. pmid: 36353484.

[21] F. P. Guengerich. "Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity". In: *Chemical Research in Toxicology* 14.6 (June 2001), pp. 611–650. DOI: 10.1021/tx0002583. pmid: 11409933.

[22] Julian P. T. Higgins and Simon G. Thompson. "Quantifying Heterogeneity in a Meta-Analysis". In: *Statistics in Medicine* 21.11 (June 15, 2002), pp. 1539–1558. DOI: 10.1002/sim.1186. pmid: 12111919.

[23] László Homolya, András Váradi, and Balázs Sarkadi. "Multidrug Resistance-Associated Proteins: Export Pumps for Conjugates with Glutathione, Glucuronate or Sulfate". In: *BioFactors (Oxford, England)* 17.1-4 (2003), pp. 103–114. DOI: 10.1002/biof.5520170111. pmid: 12897433.

[24] K. Inoue, H. Yamazaki, and T. Shimada. "Characterization of Liver Microsomal 7-Ethoxycoumarin O-deethylation and Chlorzoxazone 6-Hydroxylation Activities in Japanese and Caucasian Subjects Genotyped for CYP2E1 Gene". In: *Archives of Toxicology* 74.7 (Sept. 2000), pp. 372–378. DOI: 10.1007/s002040000151. pmid: 11043492.

[25] Hirotaka Kawakami, Sumio Ohtsuki, Junichi Kamiie, Takashi Suzuki, Takaaki Abe, and Tetsuya Terasaki. "Simultaneous Absolute Quantification of 11 Cytochrome P450 Isoforms in Human Liver Microsomes by Liquid Chromatography Tandem Mass Spectrometry with in Silico Target Peptide Selection". In: *Journal of Pharmaceutical Sciences* 100.1 (Jan. 2011), pp. 341–352. DOI: 10.1002/jps.22255. pmid: 20564338.

[26] J. König, A. T. Nies, Y. Cui, I. Leier, and D. Keppler. "Conjugate Export Pumps of the Multidrug Resistance Protein (MRP) Family: Localization, Substrate Specificity, and MRP2-mediated Drug Resistance". In: *Biochimica Et Biophysica Acta* 1461.2 (Dec. 6, 1999), pp. 377–394. DOI: 10.1016/s0005-2736(99)00169-8. pmid: 10581368.

[27] Jiabin Li, Lianqi Zhou, Huanhuan Wang, Hui Yan, Nannan Li, Rui Zhai, Fenglong Jiao, Feiran Hao, Zuyao Jin, Fang Tian, Bo Peng, Yangjun Zhang, and Xiaohong Qian. "A New Sample Preparation Method for the Absolute Quantitation of a Target Proteome Using ^{18} O Labeling Combined with Multiple Reaction Monitoring Mass Spectrometry". In: *The Analyst* 140.4 (2015), pp. 1281–1290. DOI: 10.1039/C4AN02092H. URL: http://xlink.rsc.org/?DOI=C4AN02092H (visited on 10/28/2022).

[28] Scott Michaels and Michael Zhuo Wang. "The Revised Human Liver Cytochrome P450 "Pie": Absolute Protein Quantification of CYP4F and CYP3A Enzymes Using Targeted Quantitative Proteomics". In: *Drug Metabolism and Disposition: The Biological Fate of Chemicals* 42.8 (Aug. 2014), pp. 1241–1251. DOI: 10.1124/dmd.114.058040. pmid: 24816681.

[29] Sumio Ohtsuki, Olaf Schaefer, Hirotaka Kawakami, Tae Inoue, Stephanie Liehner, Asami Saito, Naoki Ishiguro, Wataru Kishimoto, Eva Ludwig-Schwellinger, Thomas Ebner, and Tetsuya Terasaki. "Simultaneous Absolute Protein Quantification of Transporters, Cytochromes P450, and UDP-glucuronosyltransferases as a Novel Approach for the Characterization of Individual Human Liver: Comparison with mRNA Levels and Activities". In: *Drug Metabolism and Disposition: The Biological Fate of Chemicals* 40.1 (Jan. 2012), pp. 83–92. DOI: 10.1124/dmd.111.042259. pmid: 21994437.

[30] O. V. Olesen and K. Linnet. "Identification of the Human Cytochrome P450 Isoforms Mediating in Vitro N-dealkylation of Perphenazine". In: *British Journal of Clinical Pharmacology* 50.6 (Dec. 2000), pp. 563–571. DOI: 10.1046/j.1365-2125.2000.00298.x. pmid: 11136295.

[31] Kata W. Pedersen, Jakob Hansen, Jytte Banner, Jørgen B. Hasselstrøm, and Jakob R. Jornil. "Sex- and Lifestyle-Related Factors Are Associated with Altered Hepatic CYP Protein Levels in People Diagnosed with Mental Disorders". In: *Drug Metabolism and Disposition: The Biological Fate of Chemicals* 51.9 (Sept. 2023), pp. 1169–1176. DOI: 10.1124/dmd.122.001125. pmid: 37258304.

[32] Andrew Rowland, John O. Miners, and Peter I. Mackenzie. "The UDP-glucuronosyltransferases: Their Role in Drug Metabolism and Detoxification". In: *The International Journal of Biochemistry & Cell Biology* 45.6 (June 2013), pp. 1121–1132. DOI: 10.1016/j.biocel.2013.02.019. pmid: 23500526.

[33] Cathrin Seibert, Brian R. Davidson, Barry J. Fuller, Laurence H. Patterson, William J. Griffiths, and Yuqin Wang. "Multiple-Approaches to the Identification and Quantification of Cytochromes P450 in Human Liver Tissue by Mass Spectrometry". In: *Journal of Proteome Research* 8.4 (Apr. 2009), pp. 1672–1681. DOI: 10.1021/pr800795r. pmid: 19714871.

[34] J. E. Snawder and J. C. Lipscomb. "Interindividual Variance of Cytochrome P450 Forms in Human Hepatic Microsomes: Correlation of Individual Forms with Xenobiotic Metabolism and Implications in Risk Assessment". In: *Regulatory toxicology and pharmacology: RTP* 32.2 (Oct. 2000), pp. 200–209. DOI: 10.1006/rtph.2000.1424. pmid: 11067776.

[35] Stephen T. Susa, Azhar Hussain, and Charles V. Preuss. "Drug Metabolism". In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2023. pmid: 28723052. URL: http://www.ncbi.nlm.nih.gov/books/NBK442023/ (visited on 09/08/2023).

[36] Ryan H. Takahashi, William F. Forrest, Alexander D. Smith, Justine Badee, NaHong Qiu, Stephan Schmidt, Abby C. Collier, Neil Parrott, and Stephen Fowler. "Characterization of Hepatic UDP-Glucuronosyltransferase Enzyme Abundance-Activity Correlations and Population Variability Using a Proteomics Approach and Comparison with Cytochrome P450 Enzymes". In: *Drug Metabolism and Disposition: The Biological Fate of Chemicals* 49.9 (Sept. 2021), pp. 760–769. DOI: 10.1124/dmd.121.000474. pmid: 34187837.

[37] Thi Tuyet Van Tran, Hilal Tayara, and Kil To Chong. "Artificial Intelligence in Drug Metabolism and Excretion Prediction: Recent Advances, Challenges, and Future Perspectives". In: *Pharmaceutics* 15.4 (Apr. 17, 2023), p. 1260. DOI: 10.3390/pharmaceutics15041260. pmid: 37111744.

[38] Manuel Ugidos, María José Nueda, José M. Prats-Montalbán, Alberto Ferrer, Ana Conesa, and Sonia Tarazona. "MultiBaC: An R Package to Remove Batch Effects in Multi-Omic Experiments". In: *Bioinformatics (Oxford, England)* 38.9 (Apr. 28, 2022), pp. 2657–2658. DOI: 10.1093/bioinformatics/btac132. pmid: 35238331.

[39] Areti-Maria Vasilogianni, Zubida M. Al-Majdoub, Brahim Achour, Sheila Annie Peters, Jill Barber, and Amin Rostami-Hodjegan. "Quantitative Proteomics of Hepatic Drug-Metabolizing Enzymes and Transporters in Patients With Colorectal Cancer Metastasis". In: *Clinical Pharmacology and Therapeutics* 112.3 (Sept. 2022), pp. 699–710. DOI: 10.1002/cpt.2633. pmid: 35510337.

[40] Anna Vildhede, Jacek R. Wiśniewski, Agneta Norén, Maria Karlgren, and Per Artursson. "Comparative Proteomic Analysis of Human Liver Tissue and Isolated Hepatocytes with a Focus on Proteins Determining Drug Exposure". In: *Journal of Proteome Research* 14.8 (Aug. 7, 2015), pp. 3305–3314. DOI: 10.1021/acs.jproteome.5b00334. URL: https://pubs.acs.org/doi/10.1021/acs.jproteome.5b00334 (visited on 12/19/2022).

[41] Huanhuan Wang, Haifeng Zhang, Jiabin Li, Junying Wei, Rui Zhai, Bo Peng, Hailing Qiao, Yangjun Zhang, and Xiaohong Qian. "A New Calibration Curve Calculation Method for Absolute Quantification of Drug Metabolizing Enzymes in Human Liver Microsomes by Stable Isotope Dilution Mass Spectrometry". In: *Analytical Methods* 7.14 (2015), pp. 5934–5941. DOI: 10.1039/C5AY00664C. URL: http://xlink.rsc.org/?DOI=C5AY00664C (visited on 10/28/2022).

[42] Christine Wegler, Jacek R. Wiśniewski, Ida Robertsen, Hege Christensen, Jens Kristoffer Hertel, Jøran Hjelmesaeth, Rasmus Jansson-Löfmark, Anders Åsberg, Tommy B. Andersson, and Per Artursson. "Drug Disposition Protein Quantification in Matched Human Jejunum and Liver From Donors With Obesity". In: *Clinical Pharmacology and Therapeutics* 111.5 (May 2022), pp. 1142–1154. DOI: 10.1002/cpt.2558. pmid: 35158408.

[43] Ulrich M. Zanger and Matthias Schwab. "Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation". In: *Pharmacology & Therapeutics* 138.1 (Apr. 2013), pp. 103–141. DOI: 10.1016/j.pharmthera.2012.12.007. pmid: 23333322.

[44] Mingzhe Zhao, Jingsong Ma, Mo Li, Yingtian Zhang, Bixuan Jiang, Xianglong Zhao, Cong Huai, Lu Shen, Na Zhang, Lin He, and Shengying Qin. "Cytochrome P450 Enzymes and Drug Metabolism in Humans". In: *International Journal of Molecular Sciences* 22.23 (Nov. 26, 2021), p. 12808. DOI: 10.3390/ijms222312808. URL: https://www.mdpi.com/1422-0067/22/23/12808 (visited on 03/10/2023).