DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Global Proteomic Analysis of Human Liver Microsomes: Rapid Characterization and Quantification of Hepatic Drug-Metabolizing Enzymes

Brahim Achour, Hajar Al Feteisi, Francesco Lanucara, Amin Rostami-Hodjegan, and Jill Barber

Centre for Applied Pharmacokinetic Research, Division of Pharmacy and Optometry, School of Health Sciences, University of Manchester, Stopford Building, Oxford Road, Manchester, M13 9PT, United Kingdom (B.A., H.A.F, A.R-H., J.B.); Waters Corporation, Altrincham Road, Wilmslow, Milford, Cheshire East, SK9 4AX, United Kingdom (F.L.); Simcyp Limited (a Certara Company), Blades Enterprise Centre, Sheffield, S2 4SU, United Kingdom (A.R-H.).

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

RUNNING TITLE PAGE

Running Title: Global analysis of human liver microsomal sub-proteome

Corresponding Author: Dr Jill Barber

Division of Pharmacy and Optometry, School of Health Sciences, University of Manchester, Stopford Building, Oxford Road, Manchester, M13 9PT, United Kingdom

Tel: +44 (0)161 275 2369

e-mail: Jill.Barber@manchester.ac.uk

Number of text pages: 28

Number of figures: 4

Number of tables: 2

Number of references: 54

Number of words in the abstract: 246

Number of words in the introduction: 722

Number of words in the discussion: 1450

Abbreviations

ADME, absorption, distribution, metabolism and excretion; BSA, bovine serum albumin; DDI, drug-drug interaction; DME, drug-metabolizing enzyme; GO, Gene Ontology project; HLM, human liver microsomes; HPLC, high performance liquid chromatography; IMS, ion-mobility spectrometry; IVIVE, in vitro-in vivo extrapolation; LC-MS, liquid chromatography in conjunction with mass spectrometry; LysC, lysyl endopeptidase; MS/MS, tandem mass spectrometry; MALDI-TOF, matrix-assisted laser desorption ionization/time-of-flight; MPPGL, mg protein per gram liver; P450, cytochrome P450; PBPK, physiologically-based pharmacokinetics; QconCAT, quantification concatemer; UGT, uridine 5´-diphosphoglucuronosyltransferase.

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Abstract

Many genetic and environmental factors lead to inter-individual variations in metabolism and transport of drugs, profoundly affecting efficacy and toxicity. Precision dosing, targeting drug dose to a well-characterised sub-population, is dependent on quantitative models of the profiles of drug-metabolizing enzymes and transporters within that sub-population, informed by quantitative proteomics. We report the first use of ion mobility-mass spectrometry for this purpose, allowing rapid, robust, label-free quantification of human liver microsomal (HLM) proteins from distinct individuals. Approximately 1000 proteins were quantified in four samples, including an average of 75 drug-metabolizing enzymes. Technical and biological variability were distinguishable, technical variability accounting for about 10% of total variability. The biological variation between patients was clearly identified, with samples showing a range of expression profiles for cytochrome P450 and uridine 5'-diphosphoglucuronosyltransferase enzymes. Our results showed excellent agreement with previous data from targeted methods. The label-free methodology, however, allowed a fuller characterization of the in vitro system, showing, for the first time, that HLMs are significantly heterogeneous. Further, the traditional units of measurement of drug-metabolizing enzymes (pmol mg^{-1} HLM protein) are shown to introduce error arising from variability in unrelated, highly abundant proteins. Simulations of this variability suggest that up to 1.7-fold variation in apparent CYP3A4 abundance is artefactual, as are background positive correlations of up to 0.2 (Spearman correlation coefficient) between the abundances of drug-metabolizing enzymes. We suggest that protein concentrations used in pharmacokinetic predictions and scaling to in vivo clinical situations (PBPK-IVIVE) should be referenced instead to tissue mass.

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Introduction

Designing patient-specific dosage regimens within the framework of precision medicine has recently been emphasized as a key future direction in biomedical and pharmaceutical research, with physiologically-based pharmacokinetics and in vitro-in vivo extrapolation (PBPK-IVIVE) expected to play an important role in this application (Jamei, 2016). In pharmacogenomics, one of the pillars of personalized medicine, a recent survey of 517 submissions assessed by the European Medicines Agency between 1995 and 2014 showed that approximately 15% of approved medications have on-label pharmacogenomic information that directly affects therapy, indicating the recent move into tailoring drug use for specific patient sub-populations (Ehmann et al., 2015). Within this framework of targeted therapy evaluation, IVIVE-PBPK is expected to shift its focus to sub-populations with specific therapeutic needs, with increasing demand to populate these new models with expression and functional data of proteins involved in absorption, distribution, metabolism and excretion (ADME) (Turner et al., 2015; Jamei, 2016). This is supported by the substantial number of novel drug submissions (136 between 2008 and 2014) to the Food and Drug Administration for approval where PBPK has beneficially informed drug development, especially in the areas of drug-drug interactions and paediatrics (Huang et al., 2013; Jamei, 2016). Comprehensive and detailed information about the abundance and activity of ADME proteins, which play a central role in drug metabolism and disposition, is therefore required, and crucially needs to be generated with clear inter-relations with genetic, demographic, environmental and clinical information (Schadt and Björkegren, 2012; Turner et al., 2015).

Proteomics is expected to play a more prominent role in the qualitative and quantitative characterization of proteins involved in disease development and progression and modulating drug therapy, with applications ranging from biomarker discovery and disease monitoring to

dosage regimen design (Auffray et al., 2016; Masys et al., 2012). Biomolecular data acquisition and analysis should be guided by the intended clinical application with particular emphasis on disease prevention and therapy based on inter-individual variability in genetic, lifestyle and environmental factors (McGrath and Ghersi, 2016).

With recent advances in tandem mass spectrometry, many laboratories have started to contribute to the wealth of ADME protein abundance literature (Ohtsuki et al., 2012; Prasad et al., 2014; Achour et al., 2014a; Harwood et al., 2015; Vildhede et al., 2015; Fallon et al., 2016). Protein abundance values from these experiments are used in several drug pharmacokinetic prediction exercises including scaling parameters from in vitro models to in vivo clinical situations using computational PBPK models (Knights et al., 2016; Rostami-Hodjegan, 2012). However, cross-laboratory and inter-study heterogeneity highlighted recently (Achour et al., 2014b; Badée et al., 2015) have led to ongoing efforts to investigate variability originating from using different methodological workflows, taking into consideration their advantages and limitations in relation to their intended applications (Harwood et al., 2016; Al Feteisi et al., 2015).

There is little consistency in proteomic protocols used for protein quantification in a wide variety of samples, including heterogeneous membrane fractions: crude total membrane, plasma membrane and microsomal fractions (Schaefer et al., 2012; Russell et al., 2013; Groer et al., 2013; Fallon et al., 2013), and whole tissue lysates (Weiss et al., 2015; Wiśniewski et al., 2014; Wiśniewski et al., 2016a). The effects of different methodological processes on determining protein abundance were previously investigated with different levels of evidence, sometimes of conflicting nature; however, the general idea emphasized by these studies is that differences in sample preparation and in proteomic methods can contribute to considerable overall variability in end-point measurements (Balogh et al., 2013; Qiu et al., 2013; Chiva et al., 2014; Harwood et al, 2016), which makes assessment of true biological inter-individual

variability a difficult challenge. Different mass spectrometry platforms can also have an effect on the quality and robustness of analysis, with promising improvements in instrumentation making proteomic analysis more reliable. Particularly, liquid chromatography in conjunction with ion mobility spectrometry and tandem mass spectrometry (LC-IMS-MS/MS) is a relatively new approach that allows robust global analysis of entire proteomes, and has recently been applied to proteomic analysis of HeLa cell lines (Distler et al., 2014) and breast tumor xenografts (Burnum-Johnson et al., 2016).

This report describes a proof-of-concept study that aims to apply a LC-IMS-MS/MS proteomic approach to the analysis of the human liver microsomal proteome, with specific focus on quantification of the expression of drug-metabolizing enzymes. Implications of this quantitative assessment for enzyme abundance measurements and expression correlations are subsequently considered.

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Materials and Methods

Materials and chemicals

All reagents were obtained from Sigma-Aldrich (Poole, Dorset, UK) unless otherwise indicated. Lysyl endopeptidase (LysC) was purchased from Wako (Osaka, Japan) and recombinant proteomic-grade trypsin was supplied by Roche Applied Sciences (Mannheim, Germany). Label-free protein standards at 95% purity (bovine serum albumin (BSA), bovine cytochrome *c*, equine myoglobin) were purchased from Sigma-Aldrich. Solvents were of HPLC grade.

Human liver microsomal samples

Four individual human liver microsomal samples (HLM; nominally labeled as HLM01, HLM02, HLM03 and HLM04) provided by Pfizer (Groton, CT, USA), along with demographic, medication and genotype details of donors, were used in this study. Table 1 shows demographic and clinical information of the donors; suppliers of these samples were Vitron (Tucson, AZ, USA) and BD Gentest (San Jose, CA, USA). The same microsomal samples were used in the quantitative experiments using the label-free approach (the present study) and the QconCAT targeted approach (Achour et al., 2014a), which was used to analyze samples HLM01, HLM02, and HLM04. Microsomal fractions were prepared from liver tissue by the two suppliers, both using fractionation methodology based on differential centrifugation of hepatic tissue homogenates. Low speed centrifugation (10,000 *g*) was used to separate the S9 fraction (supernatant), which was followed by an ultracentrifugation step (100,000 *g*) to isolate the microsomal fraction (pellet). Ethics were covered by the suppliers.

Methodological workflows

Supplemental Figure 1 shows a summary of the label-free global proteomic workflow followed in this study. The targeted QconCAT methodology is described elsewhere (Achour et al.,

2014a). Differences between the methodological steps in these approaches are shown in Supplemental Table 1.

Proteolytic digestion of HLM samples and estimation of protein loss

Protein content in microsomal samples was determined using a colorimetric protein assay (Bradford, 1976). Proteolytic digestion and gravimetric estimation of peptide loss were carried out in triplicate using methods previously reported by Harwood et al. (2015) with slight modifications. Briefly, HLM samples (50 μg total protein mass) were suspended in ammonium bicarbonate buffer (25 mM, pH 8.0) and combined with a standard mixture of unlabeled BSA, equine myoglobin and bovine cytochrome *c* (6 μL, at 0.1, 0.02 and 0.01 mg mL^{-1}, respectively) to a final volume of 50 μL. The rationale behind using non-human standard proteins is that species-specific peptides can be found in reference proteins that should not be found in the target human proteome to allow quantification without interference due to homology in protein sequences. Mixtures were then denatured with sodium deoxycholate (acid-labile detergent) at a final concentration of 10% (w/v) for 10 min at room temperature. Disulfide bonds were reduced (dithiothreitol, 60 mM final concentration) at 56°C for 20 min and subsequently alkylated (iodoacetamide, 15 mM final concentration) in the dark at room temperature for 30 min.

Sequential enzyme proteolysis was used to increase the scope and depth of analysis and reduce the number of missed cleavages (Achour and Barber, 2013; Wiśniewski and Mann, 2012; Al-Majdoub et al 2014). Samples were diluted 1:10 with ammonium bicarbonate (25 mM) and 1 μL of LysC (1 μg μL^{-1}) was added, followed by incubation at 30ºC for 4 h. Trypsin (2.5 μL, 1 μg μL^{-1}) was then added followed by incubation at 37ºC for 18 h. After removal of detergent by acidification with trifluoroacetic acid (~pH 3.0) and centrifugation, the supernatant containing the peptides was retained and evaporated by vacuum centrifugation. Peptide loss was estimated gravimetrically as described previously (Harwood et al., 2015). Supplemental

Figure 2 shows the measured protein concentration in the HLM samples and the mass of recovered peptides following sample preparation.

MALDI-TOF mass spectrometric analysis

To confirm the quality of sample protein digests prior to LC-IMS-MS/MS analysis, digested samples were analyzed using MALDI-TOF mass spectrometry performed on an Ultraflex II instrument (Bruker, Bremen, Germany). 20 mg mL^{-1} MALDI matrix was prepared by dissolving α-cyano-4-hydroxycinnamic acid (Fluka, Buchs, Switzerland) in 0.1% trifluoroacetic acid in 50% acetonitrile in HPLC water. Samples (0.5 µL) were applied onto a MALDI target plate in triplicate. Once dry, matrix solution (0.5 µL) was added then the mixture was allowed to dry. Spectra were acquired in two *m/z* ranges: 700 to 2500 and 700 to 5000, to check for miscleaved peptides. Laser frequency of 100 Hz and intensity of 30-35% were used. Spectra of 2000 laser shots were acquired per spot. Analysis of MALD-TOF MS data was performed using FlexAnalysis version 2.2 (Bruker). Quality of spectra was checked for peptide peak intensities and *m/z* range before proceeding to LC-MS experiments.

Liquid chromatography-ion mobility spectrometry-mass spectrometry (LC-IMS-MS/MS)

Prepared HLM peptide samples were diluted 1:10, of which 2 µL were analyzed from each diluted sample. The mean HLM peptide mass analyzed in each run was 44.53 ± 5.19 ng (range: 39.59-49.66 ng). Analysis was carried out on a nanoACQUITY™ UPLC® system (Waters, Manchester, UK) connected to a SYNAPT™ G2-Si mass spectrometer (Waters). For 1D reversed-phase liquid chromatography, peptides were injected onto a Symmetry C_{18} trap column (5 µm, 180 µm × 20 mm), and then eluted onto a HSS T3 analytical column (1.8 µm, 75 µm × 250 mm), maintained at 35ºC. The LC program consisted of a gradient of 3 to 60% acetonitrile in HPLC water (acidified with 0.1% v/v formic acid) over 40 minutes with a flow rate of 300 nL min^{-1}, followed by a ramp to 95% acetonitrile for 5 min, then a return to the initial conditions over 10 min.

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Mass spectrometry was performed based on data-independent acquisition using high-definition MS^E methodology (Distler et al., 2014). The following acquisition parameters were used on the SYNAPT G2-Si: $HDMS^E$, positive electrospray (ESI+) mode, V optics, scan time 0.5 seconds, cone voltage 25 V, *m/z* range 50-2000, and lock mass [Glu^1]-Fibrinopeptide B $[M+2H]^{+2}$ 785.8426 *m/z*. Collision energy (CE) was ramped based the mobility of ions for optimal collision-induced dissociation (CID). T-Wave ion mobility (IMS) parameters were as follows: IMS T-Wave height 40 V, wave velocity 400-800 m $second^{-1}$, helium cell gas flow 180 mL min^{-1}, IMS gas flow 90 mL min^{-1}, mobility trapping release time 450 microseconds, and trap height 15 V.

Analysis of MS^E data and database searching

Analysis and searching of the LC-IMS-MS/MS data was performed using the ProteinLynx Global Server (PLGS) version 3.0.2 and $Identity^E$ (Waters) search engine, whereby the precursor ions were aligned based on retention time (RT) and drift time. Once the fragment and parent data were matched, identification was carried out by searching against a customized database containing protein sequences from human UniProt database (154,434 sequences; January 2015) and the three reference proteins. Quantification was performed using the summed intensity of the top 3 peptide ions based on the acquired label-free data for the proteins of interest and the standard proteins. The following quantification equation was applied:

$$[Protein] = [Standard] \cdot \left(\sum_{i=1}^{3} Ion\ Intensity_{protein} \Big/ \sum_{i=1}^{3} Ion\ Intensity_{standard} \right)$$

Where $[Protein]$ represents the abundance of a target protein, $[Standard]$ represents the abundance of the spiked standard in the sample (expressed in units of pmol mg^{-1} HLM protein), and the fraction refers to the ratio of the sum of the intensities of the three highest ion peaks for the target protein relative to the standard as described previously (Silva et al., 2006). The

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

integrated peak intensities of eluted peptides were used for quantification and calculations of the summed peak intensities were performed by PLGS software.

This 'top 3 CID' approach is an empirical label-free quantification method, which was previously shown to produce accurate quantification of mixtures of protein standards (Sliva et al., 2006) and to correlate with data from targeted proteomic analysis (Carroll et al., 2011). Other label-free approaches include the total protein approach, based on all quantifiable peptides from each target protein, an approach which was also previously applied to quantifying hepatic ADME proteins (Vildehede et al., 2015).

Any quantitative data below the limit of quantification were not considered reliable. The limit of quantification was nominally set using two criteria: the peptides had to be reliability identified in all three technical replicates and the replicate intensities of the peptides had to be within 20% CV of each other (i.e. consistent identification and reproducible quantification). Further appraisal of the protein standards used in this analysis is included in Supplemental Information.

Protein data annotation for function and sub-cellular localization

Proteins were classified based on their subcellular localization and function according to GO annotations (http://geneontology.org/) and database searching (http://www.uniprot.org/).

Meta-analysis of hepatic microsomal protein abundance

In order to assess the effects of variability in the most abundant ten proteins on the end-point abundance of cytochrome P450 enzymes and their expression correlations, a Matlab model was used. To inform the model with abundance values for these proteins, Medline/Pubmed (http://www.nlm.nih.gov/bsd/pmresources.html) and Web of Knowledge (http://wok.mimas.ac.uk/) electronic databases (between the years 1980 and 2016) were searched for relevant literature on the protein expression of abundant liver microsomal proteins

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

(see Table 2 for a list of these proteins) using suitable keywords including: the protein name / gene name (e.g., carboxylesterase 1 / CES1), human liver / human hepatic, protein quantification / expression / abundance, microsomes / HLM. Searches were combined and articles inspected for relevant data. Inclusion criteria were: studies that quantified primarily microsomal proteins / enzymes identified in the present analysis in adult human livers in units of, or convertible to, pmol mg^{-1} HLM protein. This analysis was used to select the ranges of the ten most highly expressed proteins in HLM samples. For the two target enzyme families (cytochrome P450 and uridine 5´-diphosphoglucuronosyltransferase enzymes), previously published meta-analyses on cytochrome P450 (Achour et al., 2014b) and UGT abundance data (Achour et al., 2014c) were used, assuming ranges and mean abundances have not changed significantly in the last two years.

Statistical data analysis and modeling

Microsoft Excel 2010 and GraphPad Prism® version 7.01 (GraphPad Software, San Diego, CA) were used for data analysis and generating graphs. Venn diagrams were generated using Venny version 2.1 (BioinfoGP, http://bioinfogp.cnb.csic.es/tools/venny/). To obtain data from graphs in publications in the meta-analysis step, GetData Graph Digitizer version 2.26 (http://www.getdata-graph-digitizer.com/) was used. The heat map was generated using QCanvas version 1.2.1 (Kim et al., 2012). Matlab R2015a (MathWorks Inc., Natick, MA, USA) was used for modeling effects of variability of the most abundant HLM proteins on abundance and correlation of P450 enzymes. Simulation was repeated ten times for n=2,000 livers in each simulation step.

Results

In this study, we set out to obtain a snapshot of the drug-metabolizing sub-proteome of four human livers, with a focus on rapid and robust sample preparation and measurement. The methodology used in this work consisted of in-solution preparation of samples followed by nanoLC-Q-IMS-TOF MS/MS, i.e. nanoflow-liquid chromatography, mass spectrometry and ion mobility both at the peptide level, then mass spectrometry at the fragment level (Supplemental Figure 1). The main aim was to identify and comprehensively quantify a complex hepatic sub-proteome in a relatively short time (<1 hour), with particular focus on drug-metabolizing enzymes.

Assessment of protein abundance measurements

The starting total protein mass for all samples was 50 µg, out of which 35.50±2.22 µg (range 33.78-38.74 µg) was recovered (Supplemental Figure 2), indicating an overall recovery of 71% as estimated gravimetrically (Harwood et al., 2015). The number of identified proteins was 901-1,018 proteins, of which 706-816 were quantifiable (Figure 1A) with abundances above the lower limit of quantification, estimated at ~0.03 fmol peptide (translating to protein abundance of ~0.6 pmol mg^{-1}).

To assess the reproducibility and precision of the methodology, the overlap of the number of quantified proteins between samples was estimated and the coefficients of variation related to technical replicates were calculated. In addition, relative error of measurements was estimated for drug-metabolizing P450 and UGT enzymes, which were quantified previously in three of the four HLM samples using QconCAT methodology (Achour et al., 2014a) (Supplemental Table 3) to allow cross-methodology comparison. The number of quantified drug-metabolizing enzymes ranged from 63 to 76, containing 10-14 drug-metabolizing P450 enzymes and 9-11 drug-metabolizing UGT enzymes (Figure 1B). Overlap of the quantified enzymes, including P450 and UGT enzymes, between the four samples is shown in Supplemental Figure 4.

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 1C shows significant linear correlation between label-free and QconCAT measurements in three samples (R^2=0.70; Rs=0.84, p<0.0001) that were analyzed previously (Achour et al., 2014a), with measurements within 2.5-fold across the two methodologies (Figure 1D). In the data of the present study, variability in cytochrome P450 and UGT enzyme abundances between samples was estimated at up to 20 fold (total inter-individual variability). Abundance values showed technical variability of less than 20% (CV) for all protein measurements. Therefore, the expected variability related to technical error; i.e., fold difference between the 5^{th} and 95^{th} centiles of measurements, calculated as $(1 + 2\,CV)/(1-2\,CV)$, was 2 fold. This means technical variability constituted up to 10% of total variability (2 fold out of a total of 20 fold). The variation due to the inherent reproducibility of mass spectrometry based experiments was therefore very small compared with the biological variability found in these samples.

Protein expression profiles of drug-metabolizing enzymes

Assessment of the protein expression levels of drug-metabolizing enzymes is summarized in Figure 2. The assessed abundances were within reported values where literature was available (Figure 2A and B). The overlap between mid-to-high abundance drug-metabolizing P450 and UGT enzymes was approximately 80%, with the most abundant enzymes being CYP3A4, CYP2E1, CYP2C9, UGT2B4 and UGT2B7. The expression profiles of the quantified drug-metabolizing enzymes in the samples under study are shown in Figure 2C, showing a distinct visual difference in the expression of enzymes in sample HLM03, which is confirmed by the heat map and rank order cluster analysis shown in Figure 2D.

Components of human liver microsomal fractions

In the liver, hepatocytes are the primary site of drug metabolism. Along with hepatocytes, liver tissue contains other non-parenchymal cell types including Kuppfer, stellate, and biliary endothelial cells. Human liver microsomes are used as an *in vitro* model of drug metabolism,

in early studies of drug development, but to date, their composition has not been systematically investigated.

A specific cell surface marker for hepatocytes, asialoglycoprotein receptor 1, ASGR1 (Peters et al., 2016), was abundant in the microsomal fraction, whereas specific markers for other types of cells were not detected in any of the samples analyzed. Within hepatocytes, the main site of metabolism is the endoplasmic reticulum; however, other sub-cellular compartments, such as the cytosol and mitochondria, also contain drug-metabolizing enzymes. Figure 3A shows identified and quantified specific membrane marker proteins that reside in the membranes of different organelles within hepatocytes (Vildhede et al., 2015). The most abundant markers were those of the endoplasmic reticulum membrane (calnexin), mitochondrial membrane (cytochrome *c* oxidase subunit 4, COX4), and plasma membrane (CD81, ATP1A1), with little difference in their abundances between analyzed samples. These specific markers suggest the presence of membranes from these compartments in the microsomal fraction and their contribution to drug metabolism in HLM preparations, although the extent of such contribution has yet to be systematically investigated.

The ten most abundant proteins in HLM samples were shown to be localized mainly in the endoplasmic reticulum (Table 2); however, when the list is expanded to include all identified proteins (1,276), the distribution of the HLM proteins was shown to be balanced between the endoplasmic reticulum (429 proteins), plasma membrane (406), cytosol/cytoplasm (411) and mitochondria (243), with overlap in a number proteins between different compartments (Figure 3B and Supplemental Figure 5A). The localization of drug-metabolizing enzymes in different cellular compartments and the corresponding overlap are also shown in Figure 3C and Supplemental Figure 5B, with most enzymes shown to be localized within the endoplasmic reticulum (50 enzymes).

Thus, although HLMs exhibit no detectable contamination from other hepatic cell types, these findings suggest HLMs are far from pure in terms of sub-cellular composition, with many sub-cellular compartments other than endoplasmic reticulum being represented.

The use of total HLM protein mass for enzyme abundance normalization

Although drug-metabolizing cytochrome P450 and UGT enzymes are mainly present in the endoplasmic reticulum, normalization of abundance values has historically been done using total HLM protein mass, routinely measured using a colorimetric assay. However, HLM samples represent a mixture of proteins from different compartments as shown above, and therefore, the effect of the most highly expressed proteins in this system, which are not directly related to drug metabolism, was investigated in this study. The top part of Table 2 shows the ten most abundant proteins in the microsomal samples, and for comparison, the bottom part shows the ranks and abundances of drug-metabolizing P450 enzymes. Figure 4A shows that although 600 proteins in the HLM fractions make up the bulk of sample mass (>99%), these top ten proteins constitute approximately 15-20% of protein mass in this fraction.

In order to assess the effect of expression variability in these ten abundant proteins on end-point measurement of drug-metabolizing cytochrome P450 enzymes, in terms of their abundance and correlation of expression, two simulations were performed. This was done based on data from literature studies, collated using meta-analysis, and our experimental data. The first simulation was intended to describe the effect of variation in the set of 10 abundant proteins on CYP3A4 abundance (Figure 4B) and the second was intended to investigate the effects of variability in these proteins on correlation between CYP3A4 and CYP2C8 (reported in the literature to be strongly correlated, $R_s=0.68$, $p<0.0001$ (Achour et al., 2014b)). This latter simulation was intended to probe how much of the strong correlation could be attributed simply to the units of measurement.

Ten simulations of 2,000 livers each with variable CYP3A4 amounts (in pmol expressed in one mg of tissue) showed that there is an overall significant decreasing trend in CYP3A4 abundance in units of pmol per mg HLM protein (Rs=-0.25 to -0.20, $p<0.0001$, n=2,000), *assuming independent regulation of expression*. When the amount of CYP3A4 in simulated livers was kept constant in tissue at the median, the apparent abundance of CYP3A4 changed 1.4-1.7-fold as a function of overall random variability in the most abundant HLM proteins (Figure 4B). When CYP2C8 and CYP3A4 were simulated independently (with variable amounts of these two enzymes in tissue), the level of correlation increased from Rs=0.0 (with no statistical significance) for random abundance values (decoy simulation) to correlation coefficients of approximately +0.1 to +0.2 ($p<0.0001$, n=2,000) in ten repeated simulations as a function of variability in the ten most abundant proteins. These preliminary simulations suggest that the variation in levels of proteins unrelated to drug metabolism can significantly influence the apparent levels of target enzymes if correction factors are not applied, such as MPPGL (mg protein per gram liver), to relate protein abundance levels to tissue mass instead of protein mass.

Discussion

Qualitative and quantitative protein characterization can afford substantial insight into the biochemical state of cells (Collins et al., 2016), and proteomics is therefore becoming increasingly important in clinical and biomedical research. Scientists and clinicians are required to make important decisions as to whether to employ a targeted approach to robustly analyze a limited set of proteins or to apply a non-targeted discovery-like methodology, which is more comprehensive but generally produces data of lower quality (Auffray et al., 2016; Collins et al., 2016). The present study involved the application of both approaches to human liver microsomal samples from the same patients to generate quantitative data for a set of drug-metabolizing enzymes, demonstrating the wide scope of analysis offered by the global (label-free) approach. It was particularly gratifying that the results showed good agreement with targeted quantification using QconCAT as a standard (Achour et al., 2015). Our results show that it is possible to obtain robust global proteomics measurements when quality control steps are taken to ensure successful implementation of quantitative analysis. In these experiments, there was rigorous quality control of sample preparation, standards, LC-IMS-MS/MS measurement and data analysis. A similar assessment of label-free quantification of a set of yeast glycolytic enzymes also demonstrated agreement with quantification using QconCAT standards (Carroll et al., 2011), further supporting previous reports of consistency in measurements carried out within the same laboratory setting (Qiu et al., 2013; Prasad and Unadkat, 2014).

The global proteomic experiment was designed to be both robust and relatively quick. The time of the experiment was intended to be less than 1 hour to demonstrate the possibility of using this technique in screening processes. For this purpose, liquid chromatography, ion mobility and mass spectrometry were used to provide three layers of separation including the physical size of analyzed peptides (Supplementary Figure 6) in order to analyze as many proteins as

possible with high reliability (Distler et al., 2014). With this snapshot type of analysis, a set of a few hundred proteins (706-816) were successfully quantified, out of which a subset of 63-76 drug and xenobiotic-metabolizing enzymes were characterized. The abundances of measured cytochrome P450 (12) and UGT (9) enzymes were within previously published literature ranges (Achour et al., 2014b; Achour et al., 2014c). The phenotypic fingerprint generated using the expression profiles and the heat map of drug-metabolizing enzymes revealed a range of abundance levels exhibiting differences between the four individual samples, with rank order cluster analysis showing sample HLM03 to have the most distinct expression profile.

The expression fingerprint of sample HLM03 showed overall lower abundances of a set of ADME proteins, exemplified by cytochrome P450 enzymes, including CYP1A2, CYP2A6, CYP2C9/19 and CYP3A4/5. Differences in the characteristics of the corresponding donor included exposure to medications, including an opioid analgesic (morphine) and a non-steroidal anti-inflammatory agent (ibuprofen), as well as certain genetic differences, including polymorphic CYP2C9 (*1/*2), CYP2C19 (*1/*2) and CYP3A5 (*3/*3). Inflammatory conditions and polymorphism were previously reported to reduce the catalytic activity of CYP1A2, CYP2C9/19 and CYP3A4/5 (Zanger and Schwab, 2013; Zanger et al., 2014). Notably, severe reduction in the expression levels of CYP3A5*3/*3 compared to the wild type and CYP3A5*1/*3 variant is well-documented in the literature (Lin et al., 2002; Achour et al., 2014a). In addition, murine hepatic expression of P450 enzymes after exposure to a derivative of morphine showed significantly lower abundances of CYP2C and CYP2E enzymes determined using immunoblotting (Sheweita, 2003). However, due to the small sample size in the present study, the effects of these differences may require further investigation in order to confirm and elucidate them.

Human liver microsomes are routinely used in the metabolic characterization of new and existing compounds, with the idea that most of the metabolic activity in these systems is

attributed to enzymes localized in the endoplasmic reticulum, which is believed to be preferentially enriched using differential centrifugation (Zhang et al., 2015). However, there is little evidence in the literature that defines the biomolecular composition of these fractions with suggestions that centrifugation can lead to either enrichment or loss of different membrane components (Harwood et al., 2014). For the purpose of addressing this gap, annotation related to subcellular localization was performed for all identified proteins in the analyzed HLM samples (1,276 proteins). This revealed information about the composition of this in vitro system, with the main components being the endoplasmic reticulum (34% of all proteins), the plasma membrane (32%) and the cytosol/cytoplasm (32%). Mitochondrial proteins also constituted a large proportion of proteins identified in HLM samples (19%). This finding is supported by the identification and quantification of specific membrane markers for the endoplasmic reticulum, mitochondria and plasma membrane in this fraction, indicating that HLM samples represent a crude, heterogeneous mixture of proteins from different cell compartments (i.e., a crude total membrane fraction), including but not limited to the endoplasmic reticulum. Technical differences in the microsomal preparation method can theoretically lead to differences in the composition of the final microsomal fraction. However, the fractionation methods used by the suppliers of these samples were very similar and the abundances of marker proteins from different cell compartments were not significantly different. Importantly, the presence of proteins from the nucleus (11%) and Golgi body (12%) shows that the initial centrifugation step may require further optimization to achieve better enrichment of endoplasmic proteins. A useful approach to eliminate the effect of fractionation on measuring protein expression profiles may be to examine the expression levels in liver tissue homogenates instead.

A similar trend was seen with annotated drug and xenobiotic-metabolizing enzymes, with most enzymes coming from the endoplasmic reticulum (nearly 60%), the cytosol and the

plasma/exosomal membrane. However, the contribution of these non-endoplasmic reticulum enzymes to drug metabolism is only hypothesized at this stage. This observation of heterogeneity is in line with the findings of a recent global proteomic analysis that showed that the distribution of drug-metabolizing enzymes in fractions of liver tissue homogenate is complex (Wiśniewski et al., 2016b). Both the current work and that of Wiśniewski et al. point to caution in applying scaling factors when enzyme abundances are measured in membrane fractions.

Implications of this level of heterogeneity in HLMs are relevant to both the way abundance levels of ADME proteins are reported and the assessment of their correlations of expression. Abundance levels of enzymes and transporters have traditionally been measured in units of pmol per mg of total microsomal protein mass. We highlight two problems with this tradition. Firstly, the total protein mass of microsomal samples represents proteins from different compartments of the cell, and the relative contribution of each compartment can, presumably, vary. In addition, the apparent expression of enzyme/transporter abundances can vary based on the total amount of protein in this system even in the cases where the level of the target enzyme/transporter is constant in tissue. In this study, the ten most abundant proteins in HLM samples were shown to constitute 15-20% of protein mass in these samples, and their expression can vary, leading to apparent variation in abundance of CYP3A4 by up to 1.7-fold ($p<0.0001$). Further, enzymes enriched in this system can achieve a level of background correlation based on variability of unrelated but highly expressed proteins, a hypothesis proposed in our earlier reports (Achour et al., 2014b; Achour et al., 2014c). This effect was simulated by randomly varying the amount of CYP2C8 and CYP3A4 in tissue and then normalizing by total protein mass with variations in the abundance of these ten unrelated proteins. This simulation revealed a level of positive background correlation (Spearman correlation coefficient, $Rs = +0.1$ to $+0.2$) with statistical significance ($p < 0.0001$) for all

assessed enzymes, further supporting the use of tissue mass, instead of total HLM protein mass, as the normalization factor, as previously advocated by Milne et al. (2011). The units of protein abundance would then be pmol mg^{-1} tissue. Although strong correlations between enzymes with common genetic regulatory mechanisms are highly expected (Wortham et al., 2007; Jover et al., 2009), the reported level of background correlation encourages exercising caution when interpreting and using weak to moderate expression relationships reported in the literature when abundance values are expressed in the traditional units even if the correlation exhibits statistical significance.

In conclusion, this report constitutes a proof-of-principle study that demonstrates the utility of snapshot global profiling of enzymes in biological systems as a screening method and raises cautionary arguments about using abundance levels of ADME proteins reported in the literature and their correlations of expression. The report also provides preliminary qualitative and quantitative details about the protein composition of HLM samples. Limitations of the current work consist of mainly the low sample size (4 HLM samples), which renders comprehensive elucidation of inter-individual variability in a population using the data in this report highly unlikely.

Acknowledgments

Professor Perdita Barran (University of Manchester) facilitated this work in many ways, and her support is gratefully acknowledged. The authors thank Professor Douglas Kell and Dr. David Ellis (University of Manchester) for allowing laboratory access, Waters Corporation (Wilmslow, Manchester, UK) for providing access to LC-IMS-MS/MS instrumentation and data analysis software, and the BioMS Core Research Facility, University of Manchester, for access to the MALDI-TOF MS instrument used in this study. Pfizer (Groton, CT) is acknowledged for providing the samples and related donor information. We would also like to thank Dr. Khaled Rabie (Manchester Metropolitan University) for assistance with simulation software and Jessica Waite and Eleanor Savill (Certara) for assistance with preparing the manuscript for submission.

Authorship Contributions

Participated in research design: Achour, Rostami-Hodjegan, Barber

Conducted experiments: Achour, Al-Feteisi, Lanucara

Performed data analysis: Achour, Lanucara

Wrote or contributed to the writing of the manuscript: Achour, Rostami-Hodjegan, Barber

References

Achour B and Barber J (2013) The activities of Achromobacter lysyl endopeptidase and Lysobacter lysyl endoproteinase as digestive enzymes for quantitative proteomics. *Rapid Commun Mass Spectrom* 27:1669–1672.

Achour B, Russell MR, Barber J and Rostami-Hodjegan A (2014a) Simultaneous quantification of the abundance of several cytochrome P450 and uridine 5′-diphospho-glucuronosyltransferase enzymes in human liver microsomes using multiplexed targeted proteomics. *Drug Metab Dispos* 42: 500–510.

Achour B, Barber J, and Rostami-Hodjegan A (2014b) Expression of hepatic drug-metabolizing cytochrome P450 enzymes and their intercorrelations: a meta-analysis. *Drug Metab Dispos* 42:1349–1356.

Achour B, Rostami-Hodjegan A and Barber J (2014c) Protein expression of various hepatic uridine 5′-diphosphoglucuronosyltransferase (UGT) enzymes and their inter-correlations: a meta-analysis. *Biopharm Drug Dispos* 34: 353–361.

Achour B, Al-Majdoub ZM, Al Feteisi H, Elmorsi Y, Rostami-Hodjegan A and Barber J (2015) Ten years of QconCATs: Application of multiplexed quantification to small medically relevant proteomes. *Int J Mass Spectrom* 391:93–104.

Al Feteisi H, Achour B, Barber J and Rostami-Hodjegan A (2015) Choice of LC-MS methods for the absolute quantification of drug-metabolizing enzymes and transporters in human tissue: a comparative cost analysis. *AAPS J* 17: 438–446.

Al-Majdoub ZM, Carroll KM, Gaskell SJ, Barber J (2014) Quantification of the proteins of the bacterial ribosome using QconCAT technology. *J Proteome Res* 13: 1211-1222.

Auffray C, Caulfield T, Griffin JL, Khoury MJ, Lupski JR, and Schwab M (2016) From genomic medicine to precision medicine: highlights of 2015. *Genome Med* 8:12.

Badée J, Achour B, Rostami-Hodjegan A and Galetin A (2015) Meta-analysis of expression of hepatic organic anion-transporting polypeptide (OATP) transporters in cellular systems relative to human liver tissue. *Drug Metab Dispos* 43:424–432.

Balogh LM, Kimoto E, Chupka J, Zhang H and Lai Y (2013) Membrane Protein Quantification by Peptide-Based Mass Spectrometry Approaches: Studies on the Organic Anion-Transporting Polypeptide Family. *J Proteomics Bioinform* 6:229–236.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254.

Burnum-Johnson KE, Nie S, Casey CP, Monroe ME, Orton DJ, Ibrahim YM, Gritsenko MA, Clauss TRW, Shukla AK, Moore RJ, Purvine SO, Shi T, Qian W, Liu T, Baker ES, Smith RD (2016) Simultaneous proteomic discovery and targeted monitoring using liquid chromatography, ion mobility spectrometry, and mass spectrometry. *Mol Cell Proteomics* 15:3694–3705.

Carroll KM, Simpson DM, Eyers CE, Knight CG, Brownridge P, Dunn WB, Winder CL, Lanthaler K, Pir P, and Malys N, et al. (2011) Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. *Mol Cell Proteomics* 10:007633.

Chiva C, Ortega M and Sabido E (2014) Influence of the digestion technique, protease, and missed cleavage peptides in protein quantitation. *J Proteome Res* 13:3979–3986.

Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger GR, Bader SL, Chan DW, Gibson BW, Gingras AC, Held JM, Hirayama-Kurogi M, et al. (2016) Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *BioRxiv* 074567.

Distler U, Kuharev J, Navarro P, Levin Y, Schild H, Tenzer S (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* 11:167–170.

Ehmann F, Caneva L, Prasad K, Paulmichl M, Maliepaard M, Llerena A, Ingelman-Sundberg M, and Papaluca-Amati M (2015) Pharmacogenomic information in drug labels: European Medicines Agency perspective. *Pharmacogenomics J* 15:201-210.

Fallon JK, Neubert H, Hyland R, Goosen TC, Smith PC (2013) Targeted quantitative proteomics for the analysis of 14 UGT1As and -2Bs in human liver using NanoUPLC-MS/MS with selected reaction monitoring. *J Proteome Res* 12: 4402–4413.

Fallon JK, Smith PC, Xia CQ, Kim MS (2016) Quantification of four efflux drug transporters in liver and kidney across species using targeted quantitative proteomics by isotope dilution nanoLC-MS/MS. *Pharm Res* 33:2280-8.

Kim N, Park H, He N, Lee HY and Yoon S (2012) QCanvas: an advanced tool for data clustering and visualization of genomics data. *Genomics Inform* 10:263–265.

Knights KM, Spencer SM, Fallon JK, Chau N, Smith PC, Miners JO (2016) Scaling factors for the in vitro-in vivo extrapolation (IV-IVE) of renal drug and xenobiotic glucuronidation clearance. *Br J Clin Pharmacol* 81:1153-64.

Lin YS, Dowling AL, Quigley SD, Farin FM, Zhang J, Lamba J, Schuetz EG, and Thummel KE (2002) Co-regulation of CYP3A4 and CYP3A5 and contribution to hepatic and intestinal midazolam metabolism. *Mol Pharmacol* 62:162–172.

Groer C, Bruck S, Lai Y, Paulick A, Busemann A, Heidecke CD, Siegmund W and Oswald S (2013) LC-MS/MS-based quantification of clinically relevant intestinal uptake and efflux transporter proteins. *J Pharm Biomed Anal* 85:253-261.

Harwood MD, Russell MR, Neuhoff S, Warhurst G and Rostami-Hodjegan A (2014) Lost in centrifugation: accounting for transporter protein losses in quantitative targeted absolute proteomics. *Drug Metab Dispos* 42:1766-1772.

Harwood MD, Achour B, Russell MR, Carlson GL, Warhurst G and Rostami-Hodjegan A (2015) Application of an LC-MS/MS method for the simultaneous quantification of human intestinal transporter proteins absolute abundance using a QconCAT technique. J Pharm Biomed Anal 110:27-33.

Harwood MD, Achour B, Neuhoff S, Russell MR, Carlson G, Warhurst G and Rostami-Hodjegan A (2016). In vitro-in vivo extrapolation scaling factors for intestinal P-glycoprotein and breast cancer resistance protein: Part I. A cross-laboratory comparison of transporter protein abundances and relative expression factors in human intestine and Caco-2 cells. *Drug Metab Dispos* 44: 297–307.

Huang SM, Abernethy DR, Wang Y, Zhao P and Zineh I (2013) The utility of modelling and simulation in drug development and regulatory review. *J Pharm Sci* 102:2912–2123.

Jamei M (2016) Recent advances in development and application of physiologically-based pharmacokinetic (PBPK) models: a transition from academic curiosity to regulatory acceptance. *Curr Pharmacol Rep* 2:161–169.

Jover R, Moya M, and Gómez-Lechón MJ (2009) Transcriptional regulation of cytochrome P450 genes by the nuclear receptor hepatocyte nuclear factor 4-alpha. *Curr Drug Metab* 10:508–519.

Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, Hoffman MA, Kohane IS and Levy HP (2012) Technical desiderata for the integration of genomic data into Electronic Health Records. *J Biomed Inform* 45:419-422.

McGrath S and Ghersi D (2016) Building towards precision medicine: empowering medical professionals for the next revolution. *BMC Med Genomics* 9:23.

Milne AM, Burchell B, and Coughtrie MWH (2011) A novel method for the immunoquantification of UDP-glucuronosyltransferases in human tissue. *Drug Metab Dispos* 39:2258–2263.

Ohtsuki S, Schaefer O, Kawakami H, Inoue T, Liehner S, Saito A, Ishiguro N, Kishimoto W, Ludwig-Schwellinger E, Ebner T and Terasaki T (2012) 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. *Drug Metab Dispos* 40:83-92.

Peters DT, Henderson CA, Warren CR, Friesen M, Xia F, Becker CE, Musunuru K and Cowan CA (2016) Asialoglycoprotein receptor 1 is a specific cell-surface marker for isolating hepatocytes derived from human pluripotent stem cells. *Development* 143:1475–1481.

Prasad B, Evers R, Gupta A, Hop CE, Salphati L, Shukla S, Ambudkar SV, and Unadkat JD (2014) Interindividual variability in hepatic organic anion-transporting polypeptides and P-glycoprotein (ABCB1) protein expression: quantification by liquid chromatography tandem mass spectroscopy and influence of genotype, age, and sex. *Drug Metab Dispos* 42:78–88.

Prasad B and Unadkat JD (2014) Comparison of heavy labeled (SIL) peptide versus SILAC protein internal standards for LC-MS/MS quantification of hepatic drug transporters. *Int J Proteomics* 2014:451510.

Qiu X, Bi YA, Balogh LM and Lai Y (2013) Absolute measurement of species differences in sodium taurocholate cotransporting polypeptide (NTCP/Ntcp) and its modulation in cultured hepatocytes. *J Pharm Sci* 102:3252-3263.

Rostami-Hodjegan A (2012) Physiologically based pharmacokinetics joined with in vitro-in vivo extrapolation of ADME: a marriage under the arch of systems pharmacology. *Clin Pharmacol Ther* 92:50-61.

Russell MR, Achour B, McKenzie EA, Lopez R, Harwood MD, Rostami-Hodjegan A and Barber J (2013) Alternative fusion protein strategies to express recalcitrant QconCAT proteins for quantitative proteomics of human drug metabolizing enzymes and transporters. *J Proteome Res* 12:5934-5942.

Schadt EE and Björkegren JL (2012) NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 4:115rv1.

Schaefer O, Ohtsuki S, Kawakami H, Inoue T, Liehner S, Saito A, Sakamoto A, Ishiguro N, Matsumaru T, Terasaki T, Ebner T (2012) Absolute quantification and differential expression of drug transporters, cytochrome P450 enzymes, and UDP-glucuronosyltransferases in cultured primary human hepatocytes. *Drug Metab Dispos* 40:93-103.

Sheweita SA (2003) Narcotic drugs change the expression of cytochrome P450 2E1 and 2C6 and other activities of carcinogen-metabolizing enzymes in the liver of male mice. *Toxicology* 191:133-142.

Silva JC, Gorenstein MV, Li GZ, Vissers JP and Geromanos SJ (2006) Absolute quantification of proteins by LCMS^{E} a virtue of parallel MS acquisition. Mol Cell Proteomics 5:144–156.

Turner RM, Park BK and Pirmohamed M (2015) Parsing interindividual drug variability: an emerging role for systems pharmacology. *Wiley Interdiscip Rev Syst Biol Med* 7:221–241.

Vildhede A, Wiśniewski JR, Noren A, Karlgren M, and Artursson P (2015) Comparative proteomic analysis of human liver tissue and isolated hepatocytes with a focus on proteins determining drug exposure. *J Proteome Res* 14:3305-3314.

Weiss F, Schnabel A, Planatscher H, van den Berg BH, Serschnitzki B, Nuessler AK, Thasler WE, Weiss TS, Reuss M, Stoll D, Templin MF, Joos TO, Marcus K, Poetz O (2015) Indirect protein quantification of drug-transforming enzymes using peptide group-specific immunoaffinity enrichment and mass spectrometry. *Sci Rep* 5:8759.

Wiśniewski JR and Mann M (2012) Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal Chem* 84:2631–2637.

Wiśniewski JR, Hein MY, Cox J and Mann M (2014) A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics* 13:3497–3506.

Wiśniewski JR, Vildhede A, Norén A, Artursson P (2016a) In-depth quantitative analysis and comparison of the human hepatocyte and hepatoma cell line HepG2 proteomes. *J Proteomics* 136:234–247.

Wiśniewski JR, Wegler C, Artursson P (2016b) Subcellular fractionation of human liver reveals limits in global proteomic quantification from isolated fractions. *Anal Biochem* 509:82-8.

Wortham M, Czerwinski M, He L, Parkinson A, and Wan Y-JY (2007) Expression of constitutive androstane receptor, hepatic nuclear factor 4 a, and P450 oxidoreductase genes determines interindividual variability in basal expression and activity of a broad scope of xenobiotic metabolism genes in the human liver. *Drug Metab Dispos* 35:1700–1710.

Zanger UM and Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Therapeut* 138:103-141.

Zanger UM, Klein K, Thomas M, Rieger JK, Tremmel R, Kandel BA, Klein M, and Magdy T (2014) Genetics, epigenetics, and regulation of drug-metabolizing cytochrome p450 enzymes. *Clin Pharmacol Ther* 95:258–261.Zhang H, Gao N, Tian X, Liu T, Fang Y, Zhou J, Wen Q, Xu B, Qi B, Gao J, and Li H (2015) Content and activity of human liver microsomal protein and prediction of individual hepatic clearance in vivo. *Sci Rep* 5:17671.

Footnotes

Financial support from the Division of Pharmacy and Optometry, School of Health Sciences, University of Manchester, is acknowledged.

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Figure legends

Figure 1 Proteomic analysis of microsomal sub-proteome in the HLM samples showing the total number of identified and quantified proteins in the fraction (A); the number of all drug-metabolizing enzymes, DME, drug-metabolizing cytochrome P450 enzymes, CYP450, and drug-metabolizing uridine 5'-diphosphoglucuronosyltransferase enzymes, UGT (B); Spearman correlation with linear regression of measurements of drug-metabolizing enzymes using the label-free methodology described in this report and measurements of the same enzymes in three of the analyzed samples using QconCAT targeted methodology (C); fold difference in abundance of enzymes of label-free measurements (method 1) in each sample relative to QconCAT measurements (method 2) expressed as a ratio ($[x_{1,i}/x_{2,i}]$ for enzyme *i*), with all pairs of measurements within approximately 2.5-fold (gray box) (D). Average fold error (AFE) is a measure of bias in the data, whereas absolute average fold error (AAFE) is a measure of scatter or spread of measurements; the closer these two measures to 1, the lower the bias and scatter in measurement; there was limited bias in the two method and a level of spread in the data (see Supplemental Table 3). In panel (C), abundances are expressed in units of pmol mg^{-1} HLM protein

Figure 2 Patterns of expression of drug-metabolizing enzymes in liver samples: cytochrome P450 enzyme abundances compared to literature values (A), UGT abundances compared to literature values (B), patterns of expression of quantified drug-metabolizing enzymes in the HLM samples (C), and heat map of the expressed P450 and UGT enzymes with samples classed using rank order clustering (D).
In panels (A) and (B), the gray highlights indicate literature derived ranges, the bars indicate literature means and the scatter points indicate experimentally derived values in this study. In panel (C), BLQ is assigned for values below the limit of quantification. In panel (D), the abundance values are normal log modified. Abundances are expressed in units of pmol mg^{-1} HLM protein

Figure 3 The abundance of specific membrane markers of hepatocytes (ASGR1), endoplasmic reticulum (calnexin), plasma membrane (ATP1A1, CD81), mitochondria (COX4) and peroxisomes (PEX14) (A), sub-cellular localization of all identified proteins (B) and drug-metabolizing enzymes (C) in analyzed samples, providing indication of the presence of membrane fractions from these organelles in human

liver microsomes. In panel (A), abundances are expressed in units of pmol mg^{-1} HLM protein. In panels (B) and (C), percentages represent the proportions of proteins identified in each sub-cellular location to the total identified number of proteins, the sum of which adds up to more than 100% due to overlap in localization of protein expression as shown in Supplemental Information

Figure 4 The contribution of the ten most abundant proteins to total HLM protein mass from 4 human livers (A) and simulated effect of variability of the top ten proteins in HLM samples on CYP3A4 abundance in 2,000 human livers (B). When the amount of CYP3A4 within tissue in simulated livers is kept constant, the abundance of CYP3A4 changes on average 1.4-1.7 fold, representing the effect of overall random variability in the most abundant proteins. Simulations were based on data obtained from this experimental study and a meta-analysis of available literature. In panel (B), abundances are expressed in units of pmol mg^{-1} HLM protein

DMD Fast Forward. Published on April 3, 2017 as DOI: 10.1124/dmd.116.074732
This article has not been copyedited and formatted. The final version may differ from this version.

Tables

Table 1 Demographic and clinical details of the individual liver donors of samples used in this study. The final column shows the suppliers of samples

Patient sample	Age (years)	Ethnicity	Gender	Cause of death	Smoking	Alcohol use	Medical history	Medication	Supplier
HLM01	31	C	F	Motor vehicle accident	Yes	No	None	None	BD Gentest
HLM02	62	C	F	Head trauma	No	No	Hypertension	Hypertension medications	BD Gentest
HLM03	41	H	F	CVA	No	Occasional	Hypertension, mild stroke	Atenolol, Dobutamine, Morphine, Nuprin	BD Gentest
HLM04	50	C	M	CVA	No	No	Healthy	None	Vitron

C, Caucasian; H, Hispanic; F, Female; M, Male; CVA, Cerebrovascular aneurysm
Human liver microsomal (HLM) samples were prepared by the suppliers using differential centrifugation of hepatic tissue homogenates

Table 2 Rank orders, abundance levels and primary subcellular localization of the ten most abundant proteins in the analyzed HLM samples and those of drug-metabolizing cytochrome P450 enzymes

	Overall rank	Protein (gene name)	Overall abundance	HLM01	HLM02	HLM03	HLM04
			Mean ± SD [a] [pmol mg^{-1}]	Mean ± SD [b] [pmol mg^{-1}] (rank)	Mean ± SD [b] [pmol mg^{-1}] (rank)	Mean ± SD [b] [pmol mg^{-1}] (rank)	Mean ± SD [b] [pmol mg^{-1}] (rank)
Top ten HLM proteins	1	Liver carboxylesterase 1 (CES1) [c]	403.14 ± 92.8	485.94 ± 24.31 (1)	396.81 ± 31.63 (1)	275.38 ± 13.45 (5)	454.42 ± 19.15 (1)
	2	Cytoplasmic actin 1 (ACTB) [d]	316.94 ± 74.63	325.51 ± 50.06 (8)	363.32 ± 21.94 (2)	370.02 ± 24.37 (2)	208.93 ± 23.54 (10)
	3	Protein disulfide isomerase (P4HB) [c]	304.25 ± 101.48	415.74 ± 12.23 (3)	324.86 ± 18.57 (3)	169.96 ± 4.02 (9)	306.44 ± 12.18 (2)
	4	78 kDa glucose-regulated protein (HSPA5) [c]	284.23 ± 62.67	352.90 ± 10.52 (6)	284.86 ± 24.83 (5)	201.29 ± 2.06 (8)	297.86 ± 10.86 (3)
	5	ATP synthase subunit beta (ATP5B) [e]	264.10 ± 96.66	140.23 ± 1.59 (24)	270.66 ± 25.62 (6)	376.41 ± 19.64 (1)	269.09 ± 2.31 (5)
	6	Protein disulfide isomerase A3 (PDIA3) [c]	262.73 ± 108.03	387.29 ± 12.61 (4)	252.14 ± 17.28 (7)	127.46 ± 2.23 (27)	286.05 ± 12.36 (4)
	7	Calreticulin (CALR) [c]	257.40 ± 86.90	372.06 ± 7.27 (5)	223.98 ± 36.44 (10)	166.32 ± 13.55 (11)	267.26 ± 4.77 (6)
	8	Haptoglobin (HP) [f]	253.47 ± 155.76	457.21 ± 34.92 (2)	292.00 ± 22.60 (4)	117.54 ± 10.34 (30)	147.11 ± 1.66 (21)
	9	Endoplasmin (HSP90B1) [c]	243.64 ± 78.95	335.09 ± 9.03 (7)	231.93 ± 17.82 (9)	144.52 ± 15.96 (14)	263.01 ± 6.13 (7)
	10	Cytochrome b5 (CYB5A) [c]	226.79 ± 18.36	251.61 ± 26.34 (11)	209.61 ± 15.40 (12)	217.01 ± 7.03 (6)	228.93 ± 36.13 (8)
Drug-metabolizing P450 enzymes [c]	44	CYP3A4	80.87 ± 58.48	126.83 ± 4.43 (26)	134.71 ± 13.65 (26)	19.32 ± 1.63 (231)	42.64 ± 1.02 (122)
	53	CYP2E1	76.28 ± 14.78	73.50 ± 6.23 (57)	74.36 ± 8.85 (65)	60.84 ± 6.38 (78)	96.43 ± 13.80 (34)
	101	CYP2C9	50.37 ± 30.63	94.26 ± 13.22 (43)	29.89 ± 0.73 (192)	28.82 ± 0.40 (171)	48.49 ± 2.83 (106)
	108	CYP4F	48.31 ± 19.96	28.37 ± 2.56 (165)	68.91 ± 8.19 (70)	34.32 ± 1.79 (142)	61.63 ± 2.33 (80)
	109	CYP2A6	48.11 ± 43.44	108.60 ± 13.01 (36)	38.44 ± 3.53 (147)	5.11 ± 0.57 (597)	40.30 ± 1.37 (127)
	180	CYP3A5	31.07 ± 9.68	37.92 ± 4.13 (117)	24.22 ± 1.54 (231)	_	_
	186	CYP1A2	30.44 ± 7.19	24.25 ± 2.59 (184)	28.72 ± 2.81 (196)	_	38.33 ± 5.88 (134)
	201	CYP2C8	27.57 ± 25.42	64.37 ± 3.03 (66)	23.22 ± 1.72 (244)	6.92 ± 0.77 (516)	15.78 ± 1.25 (296)
	262	CYP2B6	20.98 ± 1.22	20.98 ± 1.22 (225)	_	_	_
	425	CYP2D6	12.45 ± 4.84	10.74 ± 1.07 (372)	8.70 ± 1.72 (538)	17.91 ± 1.79 (245)	_
	560	CYP3A7	9.21 ± 0.62	9.21 ± 0.62 (413)	_	_	_
	815	CYP2C19	5.47 ± 0.69	5.96 ± 0.17 (542)	_	_	4.98 ± 0.76 (604)
	1066	CYP3A43	1.06 ± 0.40	0.66 ± 0.11 (711)	1.20 ± 0.12 (814)	1.04 ± 0.11 (706)	1.46 ± 0.05 (707)
	51	NADPH cytochrome P450 reductase (POR) [c]	77.95 ± 19.18	85.88 ± 3.27 (53)	100.12 ± 5.52 (35)	55.97 ± 8.61 (87)	69.85 ± 3.27 (62)

[a] SD representing combined biological and technical variability, [b] SD representing technical variability, [c] sub-cellular localization: endoplasmic reticulum, [d] subcellular localization: cytoplasm, [e] subcellular localization: mitochondria, [f] subcellular localization: secreted

Figure 1

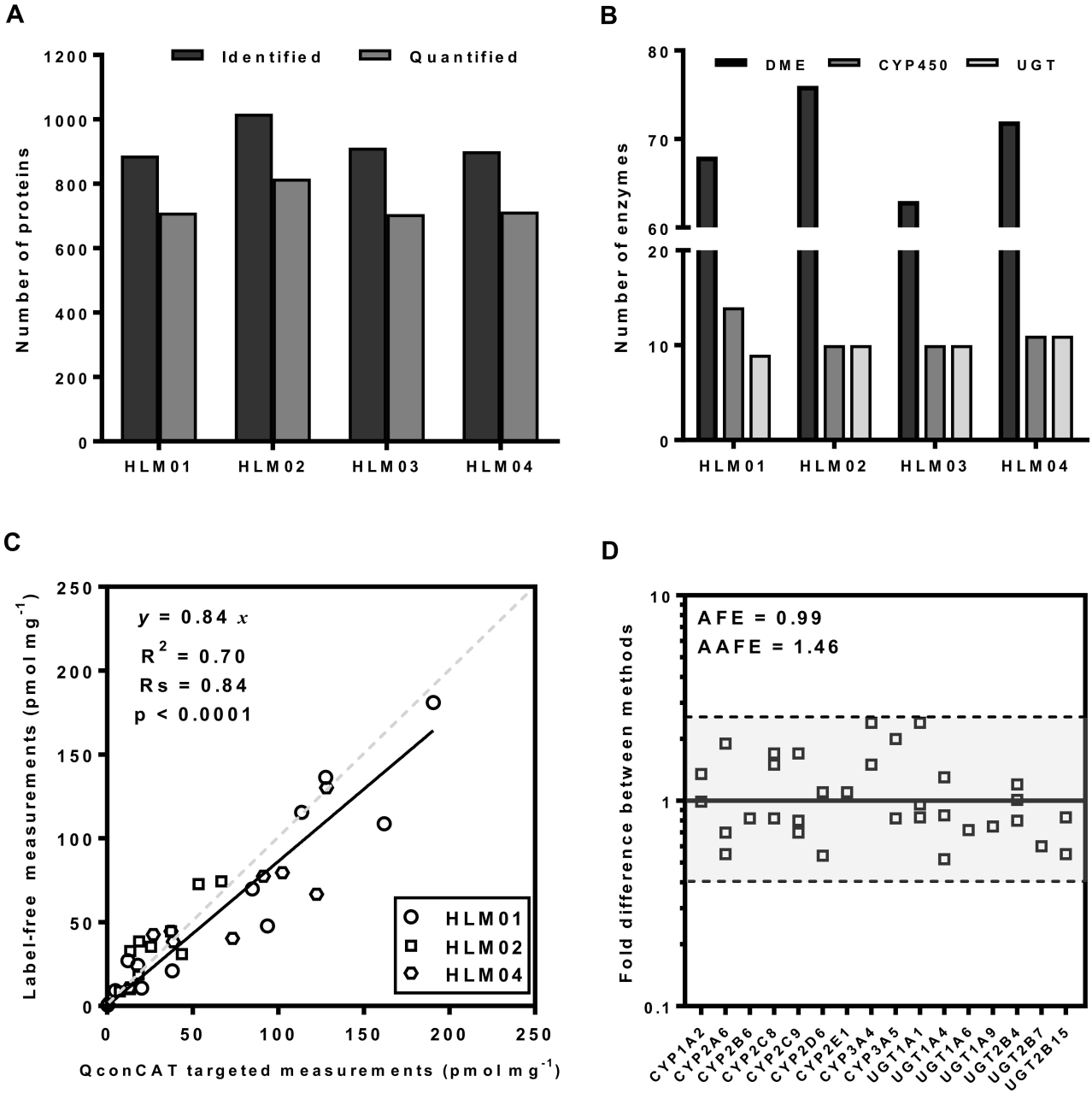

A
Identified
Quantified
Number of proteins
1200
1000
800
600
400
200
0
HLM01
HLM02
HLM03
HLM04
B
DME
CYP450
UGT
Number of enzymes
80
70
60
20
10
0
HLM01
HLM02
HLM03
HLM04
C
y = 0.84 x
R2 = 0.70
Rs = 0.84
p < 0.0001
Label-free measurements (pmol mg-1)
QconCAT targeted measurements (pmol mg-1)
HLM01
HLM02
HLM04
D
AFE = 0.99
AAFE = 1.46
Fold difference between methods
CYP1A2
CYP2A6
CYP2B6
CYP2C8
CYP2C9
CYP2D6
CYP2E1
CYP3A4
CYP3A5
UGT1A1
UGT1A4
UGT1A6
UGT1A9
UGT2B4
UGT2B7
UGT2B15

Figure 2

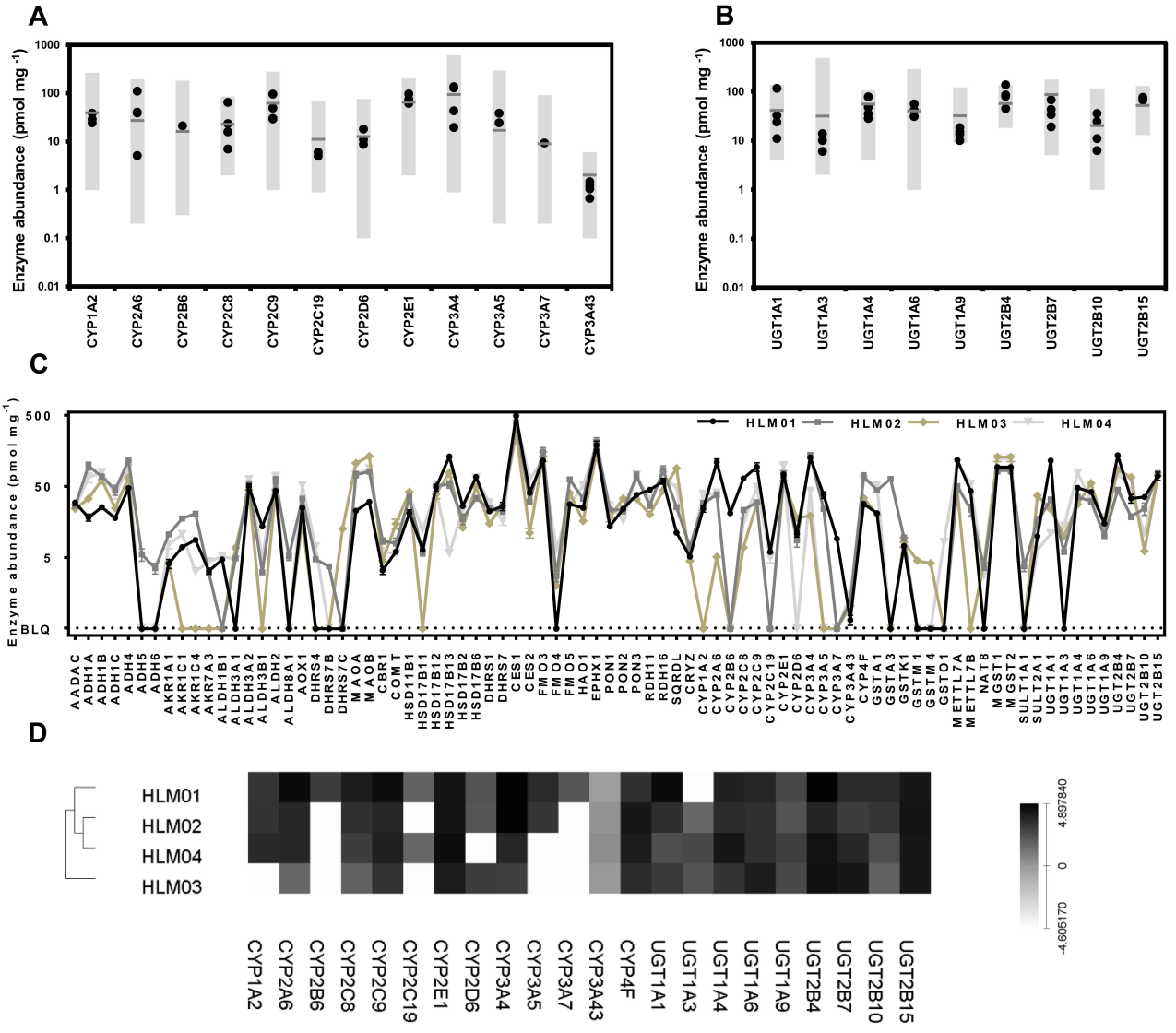

A
Enzyme abundance (pmol mg -1)
1000
100
10
1
0.1
0.01
CYP1A2
CYP2A6
CYP2B6
CYP2C8
CYP2C9
CYP2C19
CYP2D6
CYP2E1
CYP3A4
CYP3A5
CYP3A7
CYP3A43
B
Enzyme abundance (pmol mg -1)
1000
100
10
1
0.1
0.01
UGT1A1
UGT1A3
UGT1A4
UGT1A6
UGT1A9
UGT2B4
UGT2B7
UGT2B10
UGT2B15
C
Enzyme abundance (pmol mg -1)
500
50
5
BLQ
HLM01
HLM02
HLM03
HLM04
AADAC
ADH1A
ADH1B
ADH1C
ADH4
ADH5
ADH6
AKR1A1
AKR1C1
AKR1C4
AKR7A3
ALDH1B1
ALDH3A1
ALDH3A2
ALDH3B1
ALDH2
ALDH8A1
AOX1
DHRS4
DHRS7B
DHRS7C
MAOA
MAOB
CBR1
COMT
HSD11B1
HSD17B11
HSD17B12
HSD17B13
HSD17B2
HSD17B6
DHRS1
DHRS7
CES1
CES2
FMO3
FMO4
FMO5
HAO1
EPHX1
PON1
PON2
PON3
RDH11
RDH16
SQRDL
CRYZ
CYP1A2
CYP2A6
CYP2B6
CYP2C8
CYP2C9
CYP2C19
CYP2D6
CYP2E1
CYP3A4
CYP3A5
CYP3A7
CYP3A43
CYP4F
GSTA1
GSTA3
GSTK1
GSTM1
GSTM4
GSTO1
METTL7A
METTL7B
NAT8
MGST1
MGST2
SULT1A1
SULT2A1
UGT1A1
UGT1A3
UGT1A4
UGT1A6
UGT1A9
UGT2B4
UGT2B7
UGT2B10
UGT2B15
D
HLM01
HLM02
HLM04
HLM03
CYP1A2
CYP2A6
CYP2B6
CYP2C8
CYP2C9
CYP2C19
CYP2E1
CYP2D6
CYP3A4
CYP3A5
CYP3A7
CYP3A43
CYP4F
UGT1A1
UGT1A3
UGT1A4
UGT1A6
UGT1A9
UGT2B4
UGT2B7
UGT2B10
UGT2B15
4.897840
0
-4.605170

Figure 3

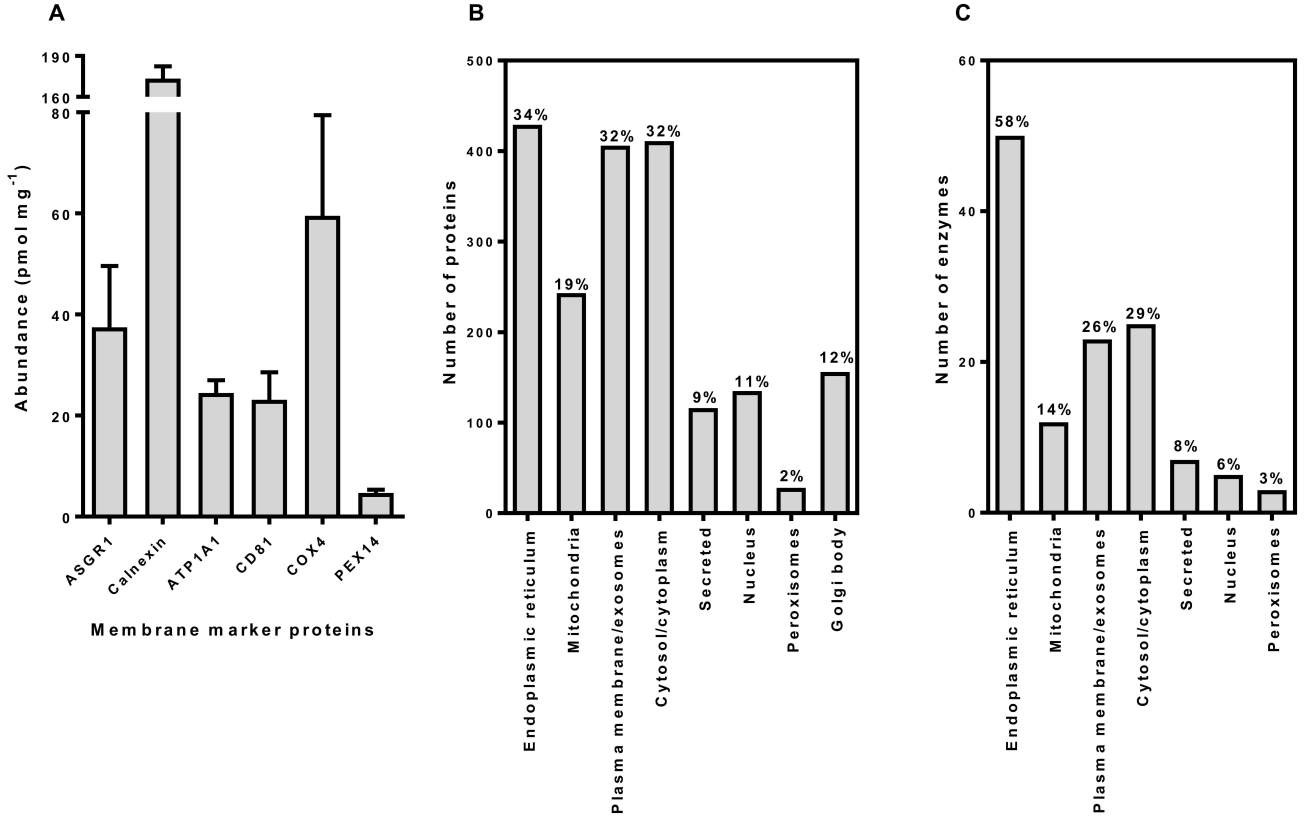
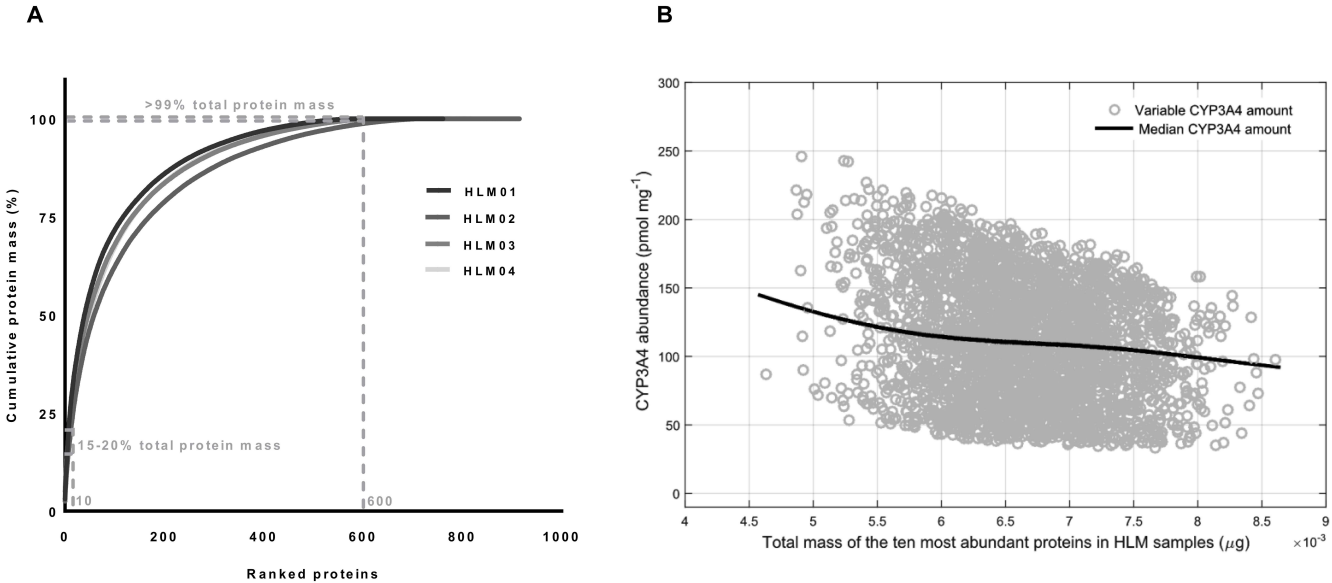

A
Abundance (pmol mg-1)
190
160
80
60
40
20
0
ASGR1
Calnexin
ATP1A1
CD81
COX4
PEX14
Membrane marker proteins
B
Number of proteins
500
400
300
200
100
0
34%
19%
32%
32%
9%
11%
2%
12%
Endoplasmic reticulum
Mitochondria
Plasma membrane/exosomes
Cytosol/cytoplasm
Secreted
Nucleus
Peroxisomes
Golgi body
C
Number of enzymes
60
40
20
0
58%
14%
26%
29%
8%
6%
3%
Endoplasmic reticulum
Mitochondria
Plasma membrane/exosomes
Cytosol/cytoplasm
Secreted
Nucleus
Peroxisomes

Figure 4


A
>99% total protein mass
100
75
50
25
0
Cumulative protein mass (%)
HLM01
HLM02
HLM03
HLM04
15-20% total protein mass
10
600
0
200
400
600
800
1000
Ranked proteins
B
300
250
200
150
100
50
0
CYP3A4 abundance (pmol mg-1)
Variable CYP3A4 amount
Median CYP3A4 amount
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
Total mass of the ten most abundant proteins in HLM samples (μg)
×10-3