Inter-individual variability in gene expression profiles in human hepatocytes and comparison with HepaRG cells

Alexandra ROGUE, Carine LAMBERT, Catherine SPIRE, Nancy CLAUDE and André GUILLOUZO

UMR INSERM U991, Faculté des Sciences Pharmaceutiques et Biologiques, Rennes, France (AR, CL, AG) ; Université de Rennes 1, Rennes, France (AR, CL, AG) ; Biologie Servier, Gidy, France (AR, CS) and Institut de Recherches Servier, Courbevoie, France (NC)
Running title: Inter-donor gene expression in human hepatocytes

Corresponding authors: Prof. Andre Guillouzo, INSERM U991, Faculté de Pharmacie, 2 avenue L. Bernard, 35043 Rennes Cedex, France; Phone: +33 2 23234391; Fax: +33 2 23 53 86; Email address: andre.guillouzo@univ-rennes1.fr or Dr. Catherine Spire, Drug Safety Assessment, Biologie Servier, 905 route de Saran, 45520 Gidy, France; Phone: +33 2 38238613; Email: address:catherine.spire@fr.netgrs.com

Number of text pages: 23
Number of tables: 2
Number of figures: 2
Number of references: 28
Number of words in the:
  – Abstract: 235
  – Introduction: 405
  – Discussion: 1337

Supplemental data: Tables: 5
Figures: 3

The abbreviations used are:
CAR, constitutive androstane receptor; CYP, cytochrome P450; DMSO, dimethylsulfoxide; FCS, fetal calf serum; FXR, farnesoid X receptor, GST, glutathione transferase; PPAR, peroxisome proliferator-activated receptor, RT-qPCR, reverse transcriptase-quantitative polymerase chain reaction; UGT, UDP-glucuronosyl transferase.
Abstract

Inter-individual variations in functions, other than drug metabolism activity, remain poorly elucidated in human liver. In the present study, the whole transcriptome of several human hepatocyte populations and the differentiated human HepaRG cell line have been analyzed and compared, using oligonucleotide pangenomic microarrays. We show that, while the variation in the percentages of expressed genes did not exceed 14% between the primary human hepatocyte populations, huge inter-individual differences in the transcript levels of many genes were observed. Variable genes were related to various functions; in addition to drug metabolism, they mainly concerned carbohydrate, amino acid and lipid metabolisms. HepaRG cells expressed from 81 to 92% of the genes active in human hepatocytes and in addition, a specific gene subset mainly related to their transformed status, some chromosomal abnormalities and the presence of primitive biliary epithelial cells. Interestingly, a relationship was evidenced between abnormal basal expression levels of some target genes and their corresponding previously reported fold changes in one out of four human hepatocyte populations treated with the hepatotoxic drug troglitazone and not with other non-hepatotoxic peroxisome proliferator-activated receptor agonists (Rogue et al., 2011). Taken altogether, our results support the view that HepaRG cells express most of the genes active in primary human hepatocytes and show that expression of most human hepatic genes can quantitatively greatly vary between individuals, thereby contributing to explain the huge inter-individual variability in susceptibility to drugs and other environmental factors.
Introduction

The liver performs major functions of the organism, which include uptake of amino acid, lipids, carbohydrate and vitamins, and their subsequent storage, conversion and release into the blood and bile. This organ is also the principal target involved in the biotransformation of chemicals with its capacity to convert hydrophobic compounds into water-soluble products that can be secreted readily from the body. A number of drugs and other xenobiotics are potentially hepatotoxic either directly or more frequently after biological activation leading to the formation of chemically reactive metabolites or generation of reactive oxygen species. Drug-induced liver injury is broadly classified into intrinsic and idiosyncratic types. While the former is dose-dependent and predictable the latter is not directly dose-dependent, unpredictable and occurs in rare patients only. Various genetic and non-genetic factors are thought to cause predisposition to idiosyncratic drug-induced liver injury that accounts for the majority of hepatotoxicity associated with medication use, probably by altering the expression level of target genes. Over 1000 drugs and herbal products have been reported to cause this type of liver injury (Biour et al., 2004; Stickel et al., 2005). There are presently no suitable preclinical models to study idiosyncratic drug-induced liver injury.

A large inter-individual variability has long been observed in the expression and corresponding activities of many genes related to drug metabolism and drug induction in either human liver or primary human hepatocytes (Morel et al., 1990; LeCluyse et al., 2000; Madan et al., 2003). These results have been confirmed and extended by analysing responses to chemicals across the entire transcriptome in primary human hepatocytes (Liguori et al., 2005; Goyak et al., 2008; Lambert et al., 2009; Rogue et al., 2011). By contrast, although a variety of genes related to other hepatic functions can be the target of hepatotoxic drugs, much less information exists about inter-individual variability in their basal expression.

In the present study we compared the whole transcriptome of six human hepatocyte populations in primary culture and human differentiated HepaRG cells. Recent studies using whole genome microarrays have indeed confirmed the great similarity between primary human hepatocytes and HepaRG cells in the responsiveness of genes related to chemical metabolism (Jennen et al., 2010; Lambert et al., 2009). We show that, whereas variations in the percentages of expressed genes were low, huge inter-individual differences in the transcripts levels were observed between the primary human hepatocyte populations and that HepaRG cells expressed most of the genes active in human hepatocytes.
Material and methods

Chemicals. Williams' E medium was supplied by Eurobio (Les Ulis, France) and fetal calf serum (FCS) by Perbio (Brebieres, France). All other chemicals were of the highest quality available.

Primary human hepatocytes. Human hepatocytes from 6 adult donors undergoing resection for primary and secondary tumors, were provided by Biopredic International (Rennes, France) (Supplemental Table 1). They were obtained by collagenase perfusion of histologically normal liver fragments and freshly seeded at a density of $17 \times 10^4$ cells/cm$^2$ in 6-well dishes in a Williams' E medium supplemented with 10% FCS, 100 units/µl penicillin, 100 µg/ml streptomycin, 1 µg/ml insulin, 2 mM glutamine and 1 µg/ml bovine serum albumin. The medium was discarded 12h after seeding and cells were thereafter maintained in serum-free medium supplemented with $10^{-7}$ M hydrocortisone hemisuccinate.

HepaRG cells. The cells were obtained from Biopredic International at passages 14 and 16. For the present studies, they were first seeded at a density of $2.6 \times 10^4$ cells/cm$^2$ in 6-well dishes in a Williams' E medium supplemented with 10% FCS, 100 units/µl penicillin, 100 µg/ml streptomycin, 5 µg/ml insulin, 2 mM glutamine and $5 \times 10^{-5}$ M hydrocortisone hemisuccinate. After two weeks of culture, they were shifted to the same culture medium supplemented with 2% dimethylsulfoxide (DMSO) for two further weeks in order to reach maximum functional activities. Media were renewed every 2–3 days. Differentiated HepaRG cell cultures were composed of both hepatocyte-like and biliary epithelial-like cells (about 50% of each type) (Cerec et al., 2007).

RNA isolation. Primary human hepatocytes (after 48 h of culture) and differentiated HepaRG cells (after 20 h of incubation in a medium deprived of DMSO) were harvested in lysis buffer (RLT buffer and β-mercaptoethanol). Total RNA was isolated using the RNeasy mini Kit (Qiagen, Venlo, Netherlands). RNA quantity and purity were assessed with a Nanodrop ND-1000 spectrophotometer (Nyxor Biotech, Paris, France) and RNA integrity was checked on a Bioanalyzer 2100 (Agilent Technologies, Massy, France).

Microarray hybridizations. Five hundred ng of total RNA from each cell culture sample were separately reverse-transcribed into double-strand cDNAs by the Moloney murine leukaemia virus reverse transcriptase and amplified for 2h at 40 °C using the Quick Amplification Labeling Kit (Agilent). The cDNAs were then transcribed into antisense cRNA.
and labelled with either CTP-Cy3 fluorescent dye for 2h at 40°C following the manufacturer’s protocol. Cyanine-labeled cRNAs were purified using RNeasy minikit (Qiagen). cRNAs were hybridized onto 4x44K Agilent Gene chip human genome Microarrays (G4112F) according to standard Agilent protocols. Human hepatocyte and HepaRG cell samples were hybridized separately but all samples of each model were hybridized simultaneously. Data analyses were performed using Rosetta Resolver v.7.0 software (Rosetta Biosoftware, Seattle, WA) for database management, quality control and analysis. All microarray data reported in this study complied with MIAME guidelines (Brazma et al., 2001). Data storage and analyses were performed using the Rosetta Resolver v.6.0 software (Rosetta Biosoftware, Seattle, WA) for database management, quality control and analysis. Ingenuity Pathways Analysis (IPA) and David analysis were used to identify relevant relationships, interactions, and pathways from normalized data or selected profiles from exploratory or statistical methods. Microarray data have been deposited in GEO (http://www.ncbi.nlm.nih.gov/geo/).

**RT-qPCR analysis.** Transcripts of some genes were also estimated by quantitative PCR in order to confirm microarrays results. Briefly, 500 ng of total RNA was reverse-transcribed into cDNA using the High-Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA). RT-qPCR was performed by the fluorescent dye SYBR Green methodology using the SYBR Green PCR Master Mix (Applied Biosystems) and the STEP one Plus (Applied Biosystems). Primer pairs for each transcript were chosen with qPrimer depot software (http://primerdepot.nci.nih.gov/). Amplification curves were read with the StepOne software V2.1 (Applied Biosystems) using the comparative cycle threshold method. The relative quantification of the steady-state mRNA levels was normalized against 18S mRNA.

**Statistical analysis and filtering** Normalization algorithms and background subtractions were automatically applied to each array to reduce systematic errors and to adjust effects due to technical rather than biological variations using Feature Extraction® and Resolver® softwares. Expressed genes (at the Entrez Gene level) were extracted from the normalized data (intensity>200 and pv<0.01). All gene sets on the arrays were included for correlation analysis. A Pearson correlation coefficient, noted r, was calculated for each hepatocyte donor versus the five other donors and for each donor versus HepaRG cells. Principal component analysis (PCA) and hierarchical clustering using the Euclidian distance associated to the ward’s min variance link heuristic criteria were performed to visualize behavior of data through the different human hepatocyte populations and HepaRG cells. One-way ANOVA (with the Benjamini & Hochberg correction, FDR=1 %, p<0.01) following by a Tukey
Kramer test (p<0.01) were performed to evidence significant differences between human hepatocytes from all the donors as well as between hepatocyte populations and HepaRG cells.
Results

Correlation coefficients between primary human hepatocytes and HepaRG cells

A total of 45520 probe sets corresponding to 20235 genes were present on each array. The presence of control probes and several probes which matched with only one gene explained these differences. The 20235 gene set was used for all transcriptomic analyses. To assess the magnitude of variation between hepatocyte cultures from the six donors across the entire transcriptome, correlation coefficients were compared and used as a global measure of similarity between donors as well as between donors and HepaRG cells. A similarity matrix was constructed for each pairwise of any two sets of the data and the Pearson correlation coefficient \((r)\) was used to represent the strength of the linear relationship between any two sets of variable (Table 1A). The correlation coefficients were relatively high and ranged from 0.88 to 0.95. The highest \(r\) value of 0.95 was observed between donors #1 and #3 and the lowest of 0.88 between donors #4 and #5 or #6. The biological replicates of HepaRG cells, corresponding to 2 different passages, which were combined together, showed close and high correlation coefficients between 0.86 and 0.88 with the 6 human hepatocyte populations. The highest \(r\) value was obtained with donors #2, #3 and #6 and the lowest with donor #1.

Hierarchical clustering and principal component analysis

All gene sets from the six human hepatocyte populations and HepaRG cells were subjected to hierarchical clustering and PCA (Figure 1). Both clustering and PCA representations showed that donors #5 and #6 were separated from the four other donors according to the principal component one on the PCA and with a distance of 0.65 on the cluster. Moreover, with a distance of 0.94, donors #3 and #4 grouped closer than with the others (Figure 1A). Both representations showed that gene expression profiles of HepaRG cells grouped separately from all human hepatocyte populations. As shown by PCA, HepaRG cells and human hepatocyte populations separated according to the principal component one whereas each human hepatocyte population separated from others according to the principal component two (Figure 1B). On the dendrogram, the distance between HepaRG cells and primary human hepatocyte populations was equal to 0.5. Noticeably, HepaRG cells clustered more closely with donors #5 and #6 than with the other donors.
Numbers of expressed genes

Only genes exhibiting an intensity signal superior to 200 associated with a p-value \(<0.01\) were considered to be expressed in primary human hepatocytes and HepaRG cells (Table 1B). Indeed, this cut-off removed most of the non-hepatic genes. Moreover, since HepaRG cells are derived from a liver tumor of a female patient, we looked for expression of genes known to be located on chromosome Y. However, numerous genes located on this chromosome are also found on other chromosomes, e.g. Interleukin 9 Receptor (IL9R) and CD99, and were, as expected, also expressed in HepaRG cells. Nevertheless, few genes were not expressed in HepaRG cells or in hepatocytes from the unique female patient (donor #6). One example was the Dead box protein 3, Y linked (DDX3Y) gene whose intensity signal was much lower than the background signal (Figure 2A).

The total number of genes expressed in human hepatocytes varied from 8335 to 9501 genes depending on the donor (Table 1B). A maximum variation of 14% was observed between donors #1 and #2. The number of genes commonly expressed between donors for pairwise comparison of any two sets of the data varied from 7842 and 9030 genes (Table 2C). Noticeably, 7780 genes were common to the 6 donors, whereas 10095 genes were expressed in at least one donor.

The number of genes expressed in HepaRG cells was analyzed at two different passages; it reached 11888 and 12036 genes at passages 14 and 16, respectively. Among them, 11691 were common (Table 1B). Comparison of the numbers of genes expressed in human hepatocytes revealed that 8168 to 9262 genes were expressed in common in at least one human hepatocyte donor and HepaRG cells, representing 81 to 92 % of the total genes expressed in at least one human hepatocyte donor (Table 1C).

Gene expression analysis across human hepatocyte donors

An ANOVA statistical analysis followed by a PostHoc Tuckey Kramer test was performed in order to identify the genes statistically differentially expressed between hepatocytes from the 6 donors (Table 1D). Their number varied between 298 and 2099. The lowest and the highest variations were observed between donors #1 and #3 and donors #1 and #6 respectively. The statistically differentially expressed genes included genes expressed at different levels between donors as well as genes expressed only in some donors. Most of them were involved in drug, carbohydrate, amino acid and lipid metabolisms; those related to metabolism of drugs and/or endogenous and other exogenous substances exhibiting the largest inter-individual
variability. Thus, genes involved in phase I drug biotransformation such as CYP2A6, CYP3A4 and CYP2E1, were markedly differentially expressed between all donors. The lowest expression values of these genes were usually found in donors #5 and #6. As an example, CYP3A4 was 31-fold less expressed in donor #6 than in donor #1. Other CYPs, such as CYP2D6, CYP1A2 and CYP2C19, were also expressed in all six donors with expression levels similarly lower in donors #5 and #6 than in others (Supplemental Table 2). Noteworthy, the prototypical nuclear receptor, the constitutive androstane receptor gene (CAR also known as NR1I3), was more expressed in donors #1, #2, #3 and #4 (Supplemental Table 2). Likewise, the CAR-dependent responsive gene, CYP2C9, was much more expressed in donors #1 to #4 than in donors #5 and #6 (Supplemental Figure 1).

Expression of many genes encoding various phase II conjugating enzymes, including several UDP-glucuronosyltransferases (UGT), such as UGT2A3 and UGT2B10, and glutathione S-transferases (GST), such as GSTA1 and GSTA2, also markedly varied from one donor to another. Moreover, GSTM1, which is known to be expressed in only 50% of the individuals, was not detected in 2 out of the 6 donors (i.e. donors #1 and #3). Various transporters were also differentially expressed across the donors; they included multidrug resistance protein (MDR) 1, multi-drug resistance associated protein (MRP) 3 and 4, and ATP binding cassette subfamily G member 5 (ABCG5). The bile salt efflux pump (BSEP also known as ABCB11) was detected in donor #1 only. Interestingly, CYP7A1 as well as ABCG5, both involved in cholesterol metabolism, were dramatically expressed in donor #1. By contrast, CYP8B1 was expressed at similar levels in all hepatocyte populations (Supplemental Table 2).

Most genes involved in other hepatic functions, such as carbohydrate and lipid metabolisms, and the urea cycle, were less differentially expressed between primary human hepatocyte populations than those related to xenobiotic metabolism. Nevertheless, few genes involved in carbohydrate [glucose-6-phosphatase (G6Pc), glycogen synthase 2 (GYS-2), hexokinase 1 (HK-1), phosphoenolpyruvate carboxykinase 2 (PCK2)], lipid [(stearoyl-CoA desaturase (SCD), CYP4A11, lipoprotein lipase (LPL), acyl-CoA dehydrogenase (ACADS)] and amino acid [4-aminobutyrate aminotransferase (ABAT), glutaminase 2 (GLS2)] metabolisms were expressed at quite variable levels across the donors (Figure 2 and Supplemental Table 2). Noticeably, most genes related to peroxisome proliferator-activated receptors (PPARs) or the farnesoid X receptor (FXR), which are nuclear receptors involved in some major hepatic functions, also showed large inter-individual variability in their expression levels, as for example apolipoprotein C3 (APOC3) which was less expressed in donors #5 and #6 (Supplemental Figure 1).
Gene expression analysis between hepatocyte donors and HepaRG cells

Venn diagrams showed that 9754 genes were expressed in at least one hepatocyte donor while 2887 genes were expressed in HepaRG cells only and 341 genes in HepaRG cells and at least one human hepatocyte population (Supplemental Figure 2). This indicates that 96% and 78% of the genes expressed in at least one human hepatocyte population and HepaRG cells respectively, were in common. These values corresponded to 7618 genes expressed in all hepatocyte populations and HepaRG cells, 162 in hepatocyte populations only and 4587 in HepaRG cells only. Many genes involved in phase I drug metabolism, such as CYP2C9, were expressed at comparable levels in both HepaRG cells and the majority of hepatocyte donors. However, a number of genes exhibited large quantitative variations in their expression levels between the two cell models. Thus, 3356 and 1532 genes were respectively at least 2-fold more or less expressed in HepaRG cells compared to all primary human hepatocyte populations. These genes usually corresponded with those extracted from the loading plot (Supplemental Figure 3 and supplemental Table 4) and responsible for the discrimination between the two cell models. A great number of genes involved in DNA repair and the cell cycle were always more expressed in HepaRG cells while genes involved in xenobiotic metabolism, complement and coagulation cascade and inflammatory processes, were frequently expressed at higher levels in primary hepatocytes. The less expressed xenobiotic-related genes in HepaRG cells included genes encoding some CYPs (e.g. CYP2D6, CYP2E1), conjugating enzymes (e.g. GSTP1, UGT1A6) and membrane transporters involved in excretion of endogenous and/or exogenous compounds (e.g. ABCG5). However, ABCB11 (BSEP) and ABCB4 (MDR3), two genes encoding bile acid transporters, were at least 1.5- and 3-fold more expressed in HepaRG cells than in human and 341 hepatocytes, with the exception of donor #1. The basolateral transporter ABCC3 (MRP3) was also much more expressed in HepaRG cells than in human hepatocytes. Less than 350 genes expressed in at least one human hepatocyte population were not detected in HepaRG cells; they included some phases I and II genes, such as CYP2C18, GSTP1 and GSTM3 (Supplemental Tables 2 and 3).

Importantly, HepaRG cells expressed 29% more genes than the human hepatocyte populations, in particular genes involved in the cell cycle, such as cyclin B1 (CCNB1), cyclin D1 (CCND1), cell division cycle (CDC) 2 and 6, as well as genes involved in focal adhesion and transcriptional activity, such as signal transducer and activator of transcription 3 (STAT3), deleted in polyposis 1 (DP1) and inhibitor of DNA binding 1 (ID1) (Figure 2).
Since these transformed cells exhibit a trisomic chromosome 7 (Gripon et al., 2002) we examined whether some of the genes located on chromosome 7 were selectively expressed. Among the 1500 genes located on this chromosome approximately 110 genes, including IL6, carnitine O-octanoyltransferase (CROT) and SLC25A40, were found to be expressed in HepaRG cells only (Table S3). Interestingly, the cytokeratin 19 (KRT19), known as a stemness and biliary cell marker, was dramatically more expressed in HepaRG cells than in human hepatocytes (Figure 2). Another biliary cell marker, the α6-integrin (Couvelard et al., 1998), was expressed in both HepaRG cells and human hepatocytes but was 2-fold higher in the former.

Is a basal gene expression level predictable of an unusual response to a hepatotoxic drug?

To date, gene expression changes induced by chemicals have been mostly focused on genes related to xenobiotic metabolism and inter-individual responses were found to be more variable than were the corresponding global basal gene expression profiles (Goyak et al., 2008; Rakhshandehroo et al., 2009; Rogue et al., 2011). Moreover, a higher induction level was frequently observed for CYPs with a low basal expression or activity (Guillouzo and Guguen-Guillouzo, 2008). We postulated that an abnormal basal expression of some target genes could lead to an abnormal response to a hepatotoxic drug in some patients and could help to explain an idiosyncratic toxicity. Four out of the six human hepatocyte populations analysed in the present study have been previously used to investigate changes in the transcriptome profiles induced by four PPAR agonists, e.g. two glitazones (troglitazone and rosiglitazone) and two glitazars (muraglitazar and tesaglitazar) (Rogue et al., 2011). Comparison of basal expression levels of target and non target genes (this study) and their PPAR-induced fold changes in the same culture conditions, showed that basal gene expression levels could greatly influence the level of response to the treatment, particularly with the hepatotoxic drug troglitazone (Table 2). For instance, pyruvate dehydrogenase kinase isozyme 4 (PDK4), an enzyme involved in fatty acid and glucose oxidation, was up- and down-regulated with 20 µM troglitazone in the two hepatocyte populations which had the lowest and the highest basal expression levels respectively. Similarly, CYP7A1, implicated in bile synthesis, as well as BSEP and ABCG5, both involved in bile acid secretion, showed 8.6-, 2.5- and 1.5-fold decrease respectively, with troglitazone, that has been reported to cause cholestasis, in donor #4 hepatocytes which had the highest gene expression levels. In addition, the heme oxygenase 1 gene (HMOX1) involved in oxidative stress and inflammation,
frequently associated with cholestasis, was also highly induced by troglitazone treatment (9.7-fold) in the same donor #4, which had the lowest basal gene expression. No such strong relationships between basal transcript levels of target genes and their drug-induced fold changes were evidenced with the three other hepatocyte populations treated with troglitazone or with the four hepatocyte populations treated either with rosiglitazone which has caused only very rare cases of hepatotoxicity or with the two glitazars which have induced only non-hepatic toxicities. No obvious relationship between basal transcript levels of target genes and their drug-induced fold changes was also observed in HepaRG cells treated with either troglitazone or the three other PPAR agonists (Table 2).

**Comparative microarray and qPCR data**
Microarray and qPCR results were compared for several genes in each hepatocyte population (Supplemental Table 5). The direction of change obtained by q-PCR was similar to that observed with microarrays for each analysis. Thus, donors #5 and #6 generally exhibited the lowest basal gene expression levels, as illustrated for CYP3A4, CYP2B6 and CAR.
Discussion

In the present study, we report the first analysis of the genome-wide expression profiles of human hepatocytes from several donors and the differentiated human HepaRG cell line. Based on a cut-off intensity signal of 200 and a technical p-value < 0.01, 7780 out of 20232 genes present on the microarrays were found to be expressed in the 6 human hepatocyte populations analysed. These genes included both annotated and non annotated genes. If most human genes were represented on the microarrays, few important ones were however lacking, e.g. UGT1A1 (Lambert et al., 2009). Over the 7780 expressed genes, 86% were found in all donors. Accordingly, in agreement with previous observations (Goyak et al., 2008), relatively high correlation coefficient values (0.88 to 0.95) were found across primary human hepatocytes from the six donors, especially between donors #1 and #3. Genes expressed in only some hepatocyte populations included the well-known polymorphic gene GSTM1 and the plasma membrane transporter BSEP. Similar observations have already been reported by measuring mRNA levels of various transporters by PCR (Jigorel et al., 2006). The choice to use human liver samples or freshly isolated human hepatocytes would probably allow to detect a few more active genes. Indeed, it is well established that when placed in culture, isolated hepatocytes exhibit early decrease in both transcriptional and translational activity of many genes, especially those encoding mature functions (Guillouzo and Guguen-Guillouzo, 2008). However, 2-day hepatocyte cultures, as used in the present study, are more representative of the conditions in which these cells are normally used in vitro.

In contrast with the low inter-individual qualitative differences not exceeding 14% in the 6 tested human hepatocyte populations, huge variations were observed in the transcript levels of a number of genes, especially among those related to drug, carbohydrate, amino-acid and lipid metabolisms. Indeed, many genes involved in xenobiotic metabolism, such as CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, GSTT1 and UGT2B7, were highly differentially expressed across individuals particularly between donors #5 and #6. It is well established that both donor characteristics (genetic polymorphism, disease, drug treatments) and culture conditions can greatly influence in vitro gene expression (Guguen-Guillouzo and Guillouzo, 2010; Russmann et al., 2010; Zhou et al., 2009). For instance, alterations of various human P450s have been reported in hepatocytes isolated from steatotic livers (Fisher et al., 2004; Gomez-Lechon et al., 2004). Noticeably, in the present study, high CAR
expression levels were found in hepatocytes from 4 out of 6 donors, perhaps because of their isolation from a drug-treated and/or a diseased liver, which could explain high expression levels of CAR-responsive genes such as CYP3A4 in these cells.

HepaRG cells expressed from 81 to 92% of the genes active in at least one human hepatocyte population with, in addition, a subset of 2887 genes. As previously reported from transcriptomic analysis of untreated and drug-treated HepaRG cells (Lambert et al., 2009; Hart et al., 2010), these cells exhibited a gene expression profile close to that of human hepatocytes, as demonstrated by hierarchical clustering, PCA and the global gene expression profile. Moreover, our data show that limited variations were evidenced when comparing different passages (number of deregulated genes, coefficient correlations, whole genome profiles) and that HepaRG cells appeared closer to certain human hepatocyte populations, especially donors #5 and #6 who exhibited low expression of several genes involved in xenobiotic (CYP3A4, CYP2C9), lipid (ACADS) and carbohydrate (PCK2) metabolisms (Figure 2 and supplemental Table 2). These data, together with the responses to treatment with several tested drugs (Lambert et al., 2009; Rogue et al., 2011), suggest that HepaRG cells behave as a primary human hepatocyte population, and could be classified as an “average human hepatocyte population” and used as such, to identify the major changes induced by a given drug at the entire transcriptome level. Accordingly, after treatment with the PPAR agonists, HepaRG cells behave as the majority of human hepatocyte donors (Rogue et al., 2011). However, it must be borne in mind that genes encoding some phase I and phase II enzymes and membrane transporters were expressed at lower levels than in primary human hepatocytes from most donors and consequently, it cannot be excluded that metabolic pathways and toxic responses obtained with some drugs could not completely reflect data obtained with primary human hepatocytes. Moreover, the levels of expression and activity of various genes related to drug metabolism are dependent upon addition of DMSO to the culture medium; they can be greatly lower in the absence of DMSO as for example shown for CYP3A4 (Aninat et al., 2006; Kanebratt and Andersson, 2008). However, whether in the presence or absence of DMSO, drug-metabolizing enzyme activities and their responsiveness to prototypical inducers remained relatively stable in differentiated HepaRG cells for a prolonged period (Antherieu et al., 2010; Josse et al., 2008; Kanebratt and Andersson, 2008).

Nearly 2900 genes were expressed solely in HepaRG cells. Many of these genes were likely related to the transformed state of these cells and included genes usually expressed in cancerous and/or stem cells, as well as genes related to the cell cycle. Indeed, at any culture
time a fraction of HepaRG cells was probably engaged in the cell cycle. In addition, HepaRG cell cultures were composed of both hepatocyte-like and bile epithelial-like cells (around 50% each), thereby explaining expression of some specific biliary cell markers such as cytokeratin 19 (Ceric et al., 2007). Moreover, HepaRG cells exhibit some karyotypic alterations consisting in a supernumerary and remodelled chromosome 7 and a translocation t(12;22) with a loss of the 12p fragment leading to a monosomy 12p (Gripon et al., 2002), which likely resulted in altered expression of a gene subset. The transformed phenotype of HepaRG cells could also lead to overexpression or repression of certain genes. Thus, as in HepG2 cells (Harris et al., 2004), STAT3, DP1 and ID1 were significantly overexpressed in HepaRG cells compared to primary human hepatocytes. Interestingly, the absence of DDX3Y, selectively located on chromosome Y, indicated the female origin of the HepaRG cell line.

Noteworthy, a number of genes not related to drug metabolism also exhibited a marked inter-individual variability. Since these are putative drug-target genes we compared basal expression levels of putative target and non target genes (this study) and corresponding fold changes induced by several PPAR agonists (Rogue et al., 2011) in the same four human hepatocyte populations and HepaRG cells. Basal expression levels of few target genes were found to greatly influence their level of response to the treatment, particularly with troglitazone, an enzyme involved in fatty acid and glucose oxidation, and CYP7A1, BSEP and ABCG5, implicated in bile acid synthesis and secretion (Table 2). Troglitazone was developed for the treatment of hyperglycemia and was withdrawn from the market for major liver damage, including cholestasis in few patients. Noteworthy, such a strong relationship between basal transcript levels and the extent of drug-induced deregulation of target genes was limited to one hepatocyte donor treated with troglitazone and not observed with the other PPAR agonists which have induced only rare non hepatic toxicities if any. These data support the view that comparative analysis of the magnitude of variation of basal and drug-induced expression of target genes in primary hepatocytes from several donors could help predicting drug-induced liver injury and that similar investigations deserve exploration with various hepatotoxic drugs.

In conclusion, the present transcriptomic study highlights the qualitative and especially quantitative inter-individual variability in gene expression profiles of human hepatocytes and the close resemblance in gene expression profiles between primary human hepatocytes and HepaRG cells. In addition, the relationship between abnormal basal expression and response of some target genes to treatment with the hepatotoxic troglitazone could reflect the variable...
susceptibility of humans to DILI and suggest that comparison of basal and drug-deregulated expression levels of target genes in primary hepatocytes from several donors deserves exploration with other idiosyncratic hepatotoxic drugs. Moreover, one may expected that the use of the new RNA and ChIP sequencing techniques will allow the elucidation of more precise information on basal and drug-induced gene expression profiles.
Aknowledgements:

We thank Dr. Wynne Ellis for careful reading of the manuscript.
Authorship Contributions.

*Participated in research design:* Rogue, Claude, Spire, Guillouzo  
*Conducted experiments:* Rogue  
*Contributed new reagents or analytic tools:* Rogue, Lambert, Spire  
*Performed data analysis:* Rogue, Lambert, Spire  
*Wrote or contributed to the writing of the manuscript:* Rogue, Claude, Spire, Guillouzo.
References


Footnote

Alexandra Rogue was a recipient of a CIFRE contract. This work was supported by the Servier Group and the EEC contract Predict-IV, number 20222. We thank the Biological Resources Centre of Rennes and Biopredic International for the supply of isolated human hepatocytes.
Legends for figures

Figure 1: Two-dimensional hierarchical clustering and principal component analysis of gene expression profiles of primary human hepatocyte populations and HepaRG cells

A: Two-dimensional hierarchical clustering of gene expression profiles of primary human hepatocyte populations and HepaRG cells.

The clustering was generated by using Resolver system software with an agglomerative algorithm Ward’s min variance link heuristic criteria and Euclidean distance metric (Intensity ≥ 200 and p ≤ 0.01). Two-dimensional clustering was performed on gene expression profiles obtained with the 6 primary human hepatocyte donors and HepaRG cells.

B: Principal component analysis of gene expression profiles in primary human hepatocyte populations and HepaRG cells.

Figure 2. Expression levels of few genes in the 6 primary human hepatocyte populations and the two HepaRG cell passages.

Intensity values of various genes involved in xenobiotic, lipid and carbohydrate metabolisms, hepatic functions and miscellaneous in primary human hepatocytes (PHH) from the 6 donors and the two passages of HepaRG cells.
Tables

Table 1: A similarity matrix of gene expression profiles of each pairwise comparison of primary human hepatocytes from 6 donors and HepaRG cells (A) and comparison of total (B), common (C) and differentially (D) expressed genes between primary human hepatocytes from 6 donors and HepaRG cells

A: Similarity matrix of gene expression profiles of each pairwise comparison of primary human hepatocytes from 6 donors and HepaRG cells

<table>
<thead>
<tr>
<th></th>
<th>Donor #2</th>
<th>Donor #3</th>
<th>Donor #4</th>
<th>Donor #5</th>
<th>Donor #6</th>
<th>HepaRG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor #1</td>
<td>0.94</td>
<td>0.95</td>
<td>0.89</td>
<td>0.93</td>
<td>0.91</td>
<td>0.86</td>
</tr>
<tr>
<td>Donor #2</td>
<td>0.93</td>
<td>0.93</td>
<td>0.89</td>
<td>0.93</td>
<td>0.93</td>
<td>0.88</td>
</tr>
<tr>
<td>Donor #3</td>
<td>0.89</td>
<td>0.92</td>
<td>0.88</td>
<td>0.92</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>Donor #4</td>
<td>0.88</td>
<td>0.88</td>
<td>0.93</td>
<td>0.93</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>Donor #5</td>
<td>0.86</td>
<td>0.88</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>Donor #6</td>
<td>0.86</td>
<td>0.88</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.88</td>
</tr>
</tbody>
</table>

The numbers in each column represent Pearson’s coefficient correlation r value.

B: Total numbers of genes expressed in the 6 human hepatocyte populations and HepaRG cell passages

<table>
<thead>
<tr>
<th>Donor or Passage number</th>
<th>Expressed genes number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor #1</td>
<td>8335</td>
</tr>
<tr>
<td>Donor #2</td>
<td>9501</td>
</tr>
<tr>
<td>Donor #3</td>
<td>9041</td>
</tr>
<tr>
<td>Donor #4</td>
<td>8918</td>
</tr>
<tr>
<td>Donor #5</td>
<td>9032</td>
</tr>
<tr>
<td>Donor #6</td>
<td>9396</td>
</tr>
<tr>
<td>HepaRG Passage 14</td>
<td>11888</td>
</tr>
<tr>
<td>HepaRG Passage 16</td>
<td>12036</td>
</tr>
</tbody>
</table>

The values in each column represent the numbers of expressed genes (intensity ≥ 200 with p<0.01).

C: Numbers of genes expressed in common between the 6 human hepatocyte populations and HepaRG cells

<table>
<thead>
<tr>
<th></th>
<th>Donor #2</th>
<th>Donor #3</th>
<th>Donor #4</th>
<th>Donor #5</th>
<th>Donor #6</th>
<th>HepaRG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor #1</td>
<td>8238</td>
<td>8239</td>
<td>8101</td>
<td>7874</td>
<td>8098</td>
<td>8168</td>
</tr>
<tr>
<td>Donor #2</td>
<td>8880</td>
<td>8796</td>
<td>8733</td>
<td>9030</td>
<td>9262</td>
<td></td>
</tr>
<tr>
<td>Donor #3</td>
<td>8670</td>
<td>8462</td>
<td>8723</td>
<td>8856</td>
<td>8721</td>
<td></td>
</tr>
<tr>
<td>Donor #4</td>
<td>8494</td>
<td>8698</td>
<td>8721</td>
<td>8856</td>
<td>9197</td>
<td></td>
</tr>
<tr>
<td>Donor #5</td>
<td>7842</td>
<td>7842</td>
<td>8698</td>
<td>8721</td>
<td>9197</td>
<td></td>
</tr>
<tr>
<td>Donor #6</td>
<td>7842</td>
<td>7842</td>
<td>8698</td>
<td>8721</td>
<td>9197</td>
<td></td>
</tr>
</tbody>
</table>

The values in each column represent the numbers of genes expressed in common between each pairwise comparison.

D: Numbers of genes differentially expressed between the 6 human hepatocyte donors and HepaRG cells

<table>
<thead>
<tr>
<th></th>
<th>Donor #2</th>
<th>Donor #3</th>
<th>Donor #4</th>
<th>Donor #5</th>
<th>Donor #6</th>
<th>HepaRG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor #1</td>
<td>933</td>
<td>298</td>
<td>541</td>
<td>1893</td>
<td>2099</td>
<td>3841</td>
</tr>
<tr>
<td>Donor #2</td>
<td>861</td>
<td>690</td>
<td>1312</td>
<td>1019</td>
<td>6376</td>
<td>6376</td>
</tr>
<tr>
<td>Donor #3</td>
<td>468</td>
<td>1914</td>
<td>1716</td>
<td>6060</td>
<td>6060</td>
<td>6060</td>
</tr>
<tr>
<td>Donor #4</td>
<td>1435</td>
<td>1322</td>
<td>6517</td>
<td>6517</td>
<td>6517</td>
<td>6517</td>
</tr>
<tr>
<td>Donor #5</td>
<td>761</td>
<td>761</td>
<td>6515</td>
<td>6517</td>
<td>6517</td>
<td>6517</td>
</tr>
<tr>
<td>Donor #6</td>
<td>6699</td>
<td>6699</td>
<td>6699</td>
<td>6699</td>
<td>6699</td>
<td>6699</td>
</tr>
</tbody>
</table>

The values in each column represent the number of differently expressed genes according to an ANOVA analysis following by a Tuckey Kramer postHoc test.
### Table 2: Comparison of basal expression and fold modulation levels of target and non target genes in 4 human hepatocyte populations and in HepaRG cells treated or not with PPAR agonists

| Gene Symbol | ACADL | ACADS | ACOT12 | ACOT2 | ACSL1 | ADH4 | ADPF | ANGPTL4 | APOA4 | APOCS | AQP3 | AQP7 | CPT1A | CPT1B | CPT2 | CROT | CRP | CYP2B6 | CYP3A4 | CYP4A11 | CYP4B1 | CYP4F3 | CYP4F22 | CYP7A1 | ELOVL6 | FABP1 | FGF21 | GK | HADHA | HMGCR | HMGCS1 | IRF7 | LPC | MBL2 | PDK4 | PEX11A | PLIN4 | PPP1R3A | PPARγ | SGK2 |
|-------------|-------|-------|--------|-------|-------|------|------|---------|-------|-------|------|------|-------|-------|------|-------|------|-------|--------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Donor #2    | 609   | 499   | 3955   | 13294 | 11507 | 1880 | 3110 | 4734   | 552   | 92070 | 2002  | 123   | 150   | 2544  | 159   | 185   | 15866 | 65866 | 329   | 2491  | 25   | 649   | 388   | 33861 | 239   | 448   | 2238  | 4409  | 10529 | 1440  | 13012 | 1461  | 779   | 338   | 338   | 444   | 197   | 431   |
| Donor #4    | 602   | 564   | 386   | 46066 | 2981  | 1227 | 56666 | 3352   | 536   | 10537 | 7232  | 114   | 172   | 1328  | 88    | 677   | 13449 | 64170 | 68    | 2156  | 13    | 1472  | 542   | 38334 | 99    | 461   | 2530  | 4550  | 11863 | 1059  | 3955  | 1167  | 215   | 274   | 348   | 564   | 297   | 405   |
| Donor #5    | 342   | 241   | 51518 | 110774 | 8491  | 595  | 10031 | 13651  | 456   | 31961 | 7200  | 148   | 163   | 1882  | 56    | 767   | 13499 | 15744 | 88    | 56666 | 13    | 510   | 542   | 17557 | 141   | 461   | 1845  | 2692  | 22255 | 11719 | 3474  | 2016  | 481   | 205   | 543   | 1040  | 289   |
| Donor #6    | 386   | 428   | 1518  | 10965  | 8619  | 812  | 1230  | 89777  | 456   | 97741 | 1720  | 128   | 1374  | 1376  | 1401 | 1401  | 15998 | 15744 | 82    | 2086  | 13    | 338   | 586   | 15755 | 1807  | 3615  | 8191  | 3599  | 12666 | 1408  | 3544  | 319   | 938   | 109    |
| Donor #7    | 33    | 35    | 1515  | 46066 | 360   | 12    | 12     | 15      | 117   | 114    | 15    | 12    | 15    | 13    | 10    | 1       | 15998 | 114    | 88    | 56    | 2028  | 10000  | 586   | 15755 | 1408  | 3615  | 8191  | 3599  | 12666 | 1408  | 3544  | 319   |
| Donor #8    | 1.0   | 2.7   | 1.2    | 1.2    | -1.1  | -4.2  | 1.6    | 1.3     | -1.3  | -1.4  | -2.4  | -1.1  | -1.7  | 1.2    | 1.1    | -2.2  | -1.1  | -1.0  | -1.1  | -1.0  | 1.5    | 1.2    | 1.3    | 1.2    | 1.1    | 1.2    |
| Donor #9    | -1.7  | -1.1  | 1.2    | 1.0    | 1.3    | -0.8  | 1.3    | 1.1     | 1.2    | 1.2    | 1.8    | 1.2    | 1.2    | 1.2    | -1.7  | 1.0    | 1.2    | 1.2    | 1.1    | 1.2    | 1.1    | 1.1    | 1.0    | 1.1    |
| Fold modulation with TRO | -1.2 | 1.1 | 1.3 | 1.1 | 1.7 | 1.9 | 2.0 | 1.9 | 1.3 | 1.3 | 1.2 | 2.9 | 1.7 | 2.9 | 1.8 | 1.2 | 1.2 | 1.2 | 1.0 | 1.2 | 1.2 | 1.0 | 1.1 |
| Fold modulation with ROSI | -1.6 | 1.1 | 1.3 | 1.2 | 2.3 | 1.8 | 1.5 | 2.6 | 1.2 | 1.2 | 1.2 | 1.1 | 1.3 | 1.1 | 1.2 | 1.1 | 1.1 | 1.1 | 1.1 | 1.0 | 1.1 | 1.2 | 1.1 |
| Fold modulation with MURA | -1.3 | 1.1 | 1.3 | 1.2 | 1.6 | 2.3 | 1.8 | 1.5 | 1.1 | 1.1 | 1.1 | 1.1 | 1.3 | 1.1 | 1.1 | 1.0 | 1.1 | 1.2 | 1.0 | 1.0 | 1.2 | 1.1 | 1.1 |
| Fold modulation with TESA | -1.3 | 1.1 | 1.4 | 1.2 | 1.9 | 2.4 | 3.0 | 1.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
2-day human hepatocyte cultures from 4 donors (#2, #4, #5 and #6) and HepaRG cells were previously treated for 24h with 20 µM of troglitazone (TRO), 50 µM of rosiglitazone (ROS), 50 µM of muraglitazar or 300 µM of tesaglitazar (TESA) (Rogue et al., 2011) and the fold changes obtained for various genes are displayed in this table. The four compounds were synthesized by the Servier Chemical Department and the two color –microarray technology was used to obtain these fold changes (Rogue et al., 2011). The same RNA samples were used in the present study for obtaining basal gene expression values (hybridizations were performed as described in the material and methods section). In bold abnormal basal expression levels and corresponding fold changes after troglitazone treatment of some genes in the human hepatocytes from donor # 4.