

DMD #31831

A comparison of whole genome gene expression profiles of HepaRG cells and HepG2 cells to primary human hepatocytes and human liver tissues

Steven N. Hart*, Ye Li*, Kaori Nakamoto, Eva-anne Subileau, David Steen, and Xiao-bo Zhong

Department of Pharmacology, Toxicology, and Therapeutics, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA (S.N.H., Y.L., K. N., X. Z.);
Biopredic International, 8-18 Rue Jean Pecker, 35000 Rennes, France (E. S., D. S.)

DMD #31831

- a) Running title: The HepaRG whole genome gene expression
- b) Correspondence author: Xiao-bo Zhong, Ph.D., Department of Pharmacology, Toxicology, and Therapeutics, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA. Telephone: 913-588-0401, Fax: 913-588-7501, E-mail: xzhong@kumc.edu.
- c) Number of text pages: 19
Number of tables: 0 (3 supplemental tables)
Number of figures: 5 (plus one supplemental figure)
Number of references: 23
Number of words in the abstract: 242
Number of words in the introduction: 738
Number of words in the discussion: 1148
- d) Non-Standard Abbreviations:
Alcohol dehydrogenase, ADH; Aldehyde dehydrogenase, ALDH; Aryl-hydrocarbon receptor, AHR; ATP-binding cassette family, ABC; Constitutive androstane receptor, CAR, NR1I3; Cytochrome P450, CYP; Flavin monooxygenase, FMO; Glutathione S-transferase, GST; *N*-acetyl transferase, NAT; Pregnane X receptor, PXR, NR1I2; Solute carrier organic anion transporter family, SLCO; Sulfotransferase, SULT; UDP glucuronosyltransferase, UGT; Xanthine dehydrogenase, XDH.

DMD #31831

ABSTRACT

HepaRG cells, derived from a female hepatocarcinoma patient, are capable of differentiating into biliary epithelial cells and hepatocytes. Importantly, differentiated HepaRG cells are able to maintain activities of many xenobiotic metabolizing enzymes and expression of the metabolizing enzyme genes can be induced by xenobiotics. The ability of these cells to express and induce xenobiotic-metabolizing enzymes is in stark contrast to the frequently used HepG2 cells. The previous studies have mainly focused on a set of selected genes; therefore, it is of significant interest to know the extent of similarity of gene expression at whole genome levels in HepaRG cells and HepG2 cells compared to primary human hepatocytes and human liver tissues. To accomplish this objective, we used Affymetrix U133 Plus 2.0 arrays to characterize the whole genome gene expression profiles in triplicate biological samples from HepG2 cells, and HepaRG cells (undifferentiated and differentiated cells), freshly isolated primary human hepatocytes, and frozen liver tissues. After using similarity matrix, principal components, and hierarchical clustering methods, we found that HepaRG cells globally transcribe genes at the levels more similar to human primary hepatocytes and human liver tissues than HepG2 cells. Particularly, many genes encoding drug processing proteins are transcribed at a more similar level in HepaRG cells than in HepG2 cells compared to primary human hepatocytes and liver samples. The transcriptomic similarity of HepaRG with primary human hepatocytes is encouraging for use of HepaRG cells in the study of xenobiotic metabolism, hepatotoxicology, and hepatocyte differentiation.

DMD #31831

Introduction

Drug-induced liver injury is one of the leading causes for the failure of drug approval and the withdrawal of approved drugs from the market (Lasser et al., 2002). Animal models are frequently used to identify potentially hazardous drugs for liver injury. However, more than 50% of drugs which induce liver injury in human clinical trials are not hepatotoxic to animals (Olson et al., 2000). Therefore, human liver cells are needed for more accurate *in vitro* screening of drug toxicity.

Freshly isolated primary human hepatocytes are currently the “gold standard” as *in vitro* human liver cells for understanding the pathways and mechanisms influencing drug metabolism and disposition as well as hepatotoxicity (LeCluyse et al., 2000; Luo et al., 2004; Kato et al., 2005). These cells, however, are fraught with difficulties, including their scarce and unpredictable availability, limited growth potential, differences in batch to batch preparation, short life-span, and propensity to undergo early and variable phenotypic alterations. CYP expression decreases quickly over time, likely due to the adaptation of cells to the culture environment (LeCluyse, 2001; Rodriguez-Antona et al., 2002). Additionally, basal gene expression in freshly isolated primary human hepatocytes is also distinctively different from one culture to another, which can introduce additional bias (Richert et al., 2006).

To overcome these difficulties, researchers have been searching for stable human liver cell lines for long time. Currently used human liver cell lines are generally derived from hepatic tumors. Unfortunately, most of them have altered gene expression profiles which lack most liver-specific functions. In particular, P450 gene expression and enzyme activities are usually very low or undetectable in these human liver cells. For example, HepG2 cells, the most frequently used human liver cell line, express many CYP genes at very low levels (Sassa et al., 1987).

DMD #31831

Although some CYP genes, such as CYP1A1 and CYP3A7, are expressed in HepG2 cells (Ogino et al., 2002), these P450 members are fetal-specific and not expressed in most adult livers. These facts indicate that many changes in gene expression have happened in HepG2 cells after they were derived from the liver tissue of a differentiated hepatocellular carcinoma.

Recently, a new human liver cell line, HepaRG, is available. Although this cell line is also derived from a female hepatocarcinoma patient, unlike other human liver cell lines, HepaRG cells express many drug processing genes at similar levels compared to primary human hepatocytes under a certain culture condition (Aninat et al., 2006; Kanebratt and Andersson, 2008a). These drug processing genes encode Phase-I drug metabolizing enzymes (CYP1A2, 2B6, 2C9, 2E1, and 3A4), Phase-II enzymes (UDP glucuronosyltransferase 1 family, polypeptide A1, UGT1A1; glutathione S-transferase alpha 1, GSTA1; GSTA4, and GSTM1), gene regulatory proteins (aryl-hydrocarbon receptor, AHR; pregnane X receptor, PXR; constitutive androstane receptor, CAR), liver-specific proteins (albumin, haptoglobin, and aldolase B), as well as alpha-fetoprotein, glutathione-related enzymes (γ -glutamylcysteine synthase regulatory subunit, γ -glutamylcysteine synthase catalytic subunit, glutathione synthase, and glutathione reductase), and thioredoxin. The activities of several Phase-I and Phase-II drug metabolizing enzymes were also comparable between HepaRG and freshly isolated human hepatocytes (Aninat et al., 2006). HepaRG cells also respond to PXR, CAR, and AhR activators, resulting in induction of CYP1A1, CYP1A2, CYP2B6, CYP2C8, CYP2C9, CYP2C19, and CYP3A4 *in vitro* (Kanebratt and Andersson, 2008b; Lambert et al., 2009a; Lambert et al., 2009b).

HepaRG cells can maintain a proliferative status in an undifferentiated culture media for several weeks at sub-confluence. At confluence, and with the addition of a differentiation-inducing culture medium, HepaRG cells are capable of differentiating into biliary epithelial cells

DMD #31831

and hepatocytes (Gripon et al., 2002). The genes encoding liver-specific factors, drug-metabolizing enzymes, transporters, and transcription factors are stably expressed over a multi-week culture period. Given the stable expression of these liver enriched factors over a long time in culture and the activity of several drug-metabolizing enzymes, HepaRG cells have been touted as surrogates to primary human hepatocytes for drug metabolism and disposition studies (Guillouzo et al., 2007; Hewitt et al., 2007).

The ability of the HepaRG cells to express and induce xenobiotic-metabolizing enzymes is in stark contrast to the frequently used human liver cell line, HepG2. Although HepaRG cells and HepG2 cells have been compared to human primary hepatocytes and liver tissues for their gene expression and enzyme activities of drug metabolism, the studies were done in a limited set of genes. Therefore, it is of significant interest to know the extent of similarity of gene expression at whole genome levels of HepaRG cells and HepG2 cells compared to primary human hepatocytes and human liver tissues.

DMD #31831

Material and Methods

Human liver tissues, primary human hepatocytes, and cultured cells

Human liver tissues. Three different human liver tissue samples were provided by XenoTech, LLC (Lenexa, Kansas). These subjects were Caucasian females at ages of 42, 56, and 60, respectively. The female samples were chosen because HepaRG cell line was derived from a female patient. The samples were acquired by XenoTech through the Midwest Transplant Network (Westwood, Kansas). The livers were cooled immediately after procurement with a cold perfusion solution and frozen within 6 hours.

Primary human hepatocytes. Three different primary human hepatocyte samples were provided by Biopredic International (Rennes, France). The primary human hepatocytes were isolated from the liver tissues donated by three Caucasian female patients undergoing resection for primary or secondary tumors at ages of 54, 65, and 76, respectively. The hepatocytes were isolated by collagenase perfusion of histologically normal liver fragments and seeded overnight hepatocyte monolayers in seeding medium. After two days culture in short term culture medium, total RNA was isolated from hepatocytes monolayer with Trizol. All liver fragments were not infected by hepatitis B, hepatitis C, and HIV1 viruses.

HepaRG cells. HepaRG cells were obtained from Biopredic International (Rennes, France). The cells in the original culture dish were detached by gentle trypsinization and seeded at 1×10^5 undifferentiated cells/cm² (high density) in hepatocyte wash medium (Invitrogen Corporation, Carlsbad, CA) supplemented with additives for growth media (Biopredic International, Rennes, France). The cells were incubated at 37°C and 5% CO₂. The medium was renewed every 3 days. After incubation for 14 days, the undifferentiated HepaRG cells were induced to

DMD #31831

differentiate with additives for differentiation media (Biopredic International, Rennes, France) for another 14 days. That medium was also renewed every 3 days.

HepG2 cells. HepG2 cells were obtained from the ATCC Cell Biology Collection (Manassas, Virginia). The cells were seeded at 1×10^5 cells/cm² in 75 cm² flask with a minimum essential medium supplemented with 10% FBS (Sigma, St. Louis, Missouri) and 5% of penicillin/streptomycin (Sigma, St. Louis, Missouri). The cells were incubated at 37°C and 5% CO₂. The cells were used at the time of 90-100% confluence.

Total RNA isolation

Total RNAs of the liver tissues, primary human hepatocytes, undifferentiated and differentiated HepaRG cells, and HepG2 cells were isolated by using Invitrogen TRIZOL reagent (Invitrogen, Carlsbad, California) following the manufacturer's instructions. RNA quality, quantity, and integrity were analyzed by utilizing the Agilent Bioanalyzer 2100 (Agilent Technologies, Amstelveen, The Netherlands).

Whole genome gene expression

The whole genome gene expression profiles of the adult liver tissues, primary human hepatocytes, differentiated and undifferentiated HepaRG cells, and HepG2 cells were determined by the Microarray Core Facility at the University of Kansas Medical Center using Affymetrix U133 Plus 2.0 arrays (Affymetrix, Santa Clara, California) in triplicate biological replicates of each sample type. The target preparation, library labeling, hybridization, post wash, and signal scanning were performed based on the Affymetrix manufacturer's instructions.

Microarray data analysis.

The raw microarray data in CEL files are available in the Gene Expression Omnibus with accession number GSE18269 at <http://www.ncbi.nlm.nih.gov/geo/>. The microarray data were

DMD #31831

normalized using GC-RMA (GeneChip Robust Multichip Average) algorithm (Wu et al., 2004) implemented in the R package *affy* (Wettenhall et al., 2006) when any two or more data sets are compared. “Present”, “Marginal” and “Absent” calls were made in R using the MAS5 algorithm in the *affy* package (Irizarry et al., 2003). A linear model was used to average data among three replicate arrays and also look for variability among them. A probe was removed if it did not correspond to a mapped gene or not register at least 2 “Present” calls in triplicate data sets by the MAS5 algorithm for all five groups. The remaining probes, hereafter defined as quality filtered probes, were used for further analysis. The similarity of whole genome gene expression profiles of HepaRG cells and HepG2 cells compared to human liver tissue and primary human hepatocytes was analyzed by similarity matrix, principal components, and hierarchical clustering. The similarity matrix between any two sets of the data was presented by the Pearson product-moment correlation coefficient (r) value which measured the strength of the linear relationship between two sets of variables. Principal component analysis was applied to identify similarity and differences of the whole genome gene expression profiles among the different samples. In addition to the whole genome gene expression profiles, some liver specific functional pathways, particularly the drug processing pathways were also compared among the different samples.

Pathway analysis

To define significant pathway differences during differentiation of HepaRG cells, we used the Functional Annotation Clustering Tool in DAVID (Glynn et al., 2003). Gene lists were made of Affymetrix IDs where the average replicate difference was greater or less than a \log_2 value of 1. Each gene list was uploaded using Affymetrix IDs and run against a background containing only the quality filtered probe sets. The Group Enrichment Score, which represents the

DMD #31831

geometric mean (in \log_2 scale) of member's p -values in a corresponding annotation cluster, was used to rank biological significance. Thus, the top ranked annotation groups most likely have consistent lower p -values for their annotation members. For a pathway or process to be defined, the Enrichment Score was set at 2.

DMD #31831

Results

Whole genome gene expression profiles of HepG2 cells, undifferentiated and differentiated HepaRG cells, primary human hepatocytes, and human liver tissues were generated by using Affymetrix U133 Plus 2.0 arrays in triplicate samples. A total of 54,675 probe sets existed on each array. After the probes which did not correspond to a mapped gene or not register at least 2 “Present” calls in the triplicate data sets for all five groups were removed, the remaining quality filtered 30,849 probe sets were selected for similarity and difference analysis by similarity matrix, principal components, and hierarchical clustering methods. Hybridization signal intensities in log2 scale on the 30,849 probe sets among the tested samples were presented in Supplemental Table S1.

Similarity matrix analysis. A similarity matrix was constructed for each pairwise comparison of any two sets of the data (**Fig. 1A**). The Pearson product-moment correlation coefficient (r) was used to represent the strength of the linear relationship between any two sets of variables. The relative higher r values (0.949-0.996) were found between any two replicates in each type of the five groups with a range between 0.991-0.995 in HepG2 cells, 0.995-0.996 in undifferentiated HepaRG cells, 0.989-0.995 in differentiated HepaRG cells, 0.979-0.984 in primary human hepatocytes, and 0.949-0.971 in human livers. The relative lower r values (0.768-0.937) were observed between any two sets of the data from the different types of samples. The r values in each group were then averaged to represent similarity of whole genome gene expression between any two groups of the samples (**Fig. 1B**). The highest r value (0.966) was found between undifferentiated and differentiated HepaRG cells. The second highest r value (0.920) existed between primary human hepatocytes and human livers. The r value between human primary hepatocytes and undifferentiated HepaRG (0.887) or between human

DMD #31831

primary hepatocytes and differentiated HepaRG (0.891) was higher than the r value between human primary hepatocytes and HepG2 (0.813). Similarly, the r value between human livers and undifferentiated HepaRG (0.883) or between human livers and differentiated HepaRG (0.881) was higher than the r value between human livers and HepG2 (0.791).

We further examined the number of probes with signal intensities different by more than 2 fold between any two sets of the data. **Fig. 2** shows that about 10% of the total quality filtered 30,849 probe sets expressed differently between undifferentiated and differentiated HepaRG cells. The number of probes increased to approximate 22% between human liver tissues and primary human hepatocytes. The differentially expressed probes were 26-28% between HepaRG cells and human liver tissues or primary human hepatocytes. However, up to about 37-39% of the probes were differently expressed between HepG2 cells and human liver tissues or primary human hepatocytes or HepaRG cells.

We further characterized the 10% of probes with expression levels different by more than 2 fold between undifferentiated and differentiated HepaRG cells. The probes were listed in Supplemental Table S2. A total of 1321 probes had 2 fold higher signal intensities in differentiated HepaRG cells than in undifferentiated HepaRG cells. Pathways analysis indicated that the up-regulated genes included xenobiotic and steroid metabolism, cell cycle genes, DNA replication and repair, and nuclear and ER proteins. Another 1831 probes had 2 fold lower signal intensities in differentiated HepaRG cells than in undifferentiated HepaRG cells. The down-regulated genes during HepaRG differentiation were involved in developmental processes, extracellular signaling, actin binding, and amino acid metabolism.

Principal components analysis. The similarity and differences in whole genome gene expression among HepG2 cells, undifferentiated and differentiated HepaRG cells, primary

DMD #31831

human hepatocytes, and human liver tissues were further highlighted by Principal Components Analysis (PCA). The intensities of the quality filtered 30,849 probe sets were first log₂ transformed. Three replicate sets of the data were averaged and then used in the PCA analysis. Three most variable principal components (PC1, PC2, and PC3) were plotted in three dimensions in **Fig. 3**. HepG2, undifferentiated and differentiated HepaRG, primary human hepatocytes, and human livers contributed nearly equal to the variations in PC1 (93.0%). HepG2 made a major contribution to the variations in PC2 (3.3%). Undifferentiated and differentiated HepaRG cells contributed the majority of the variations in PC3 which only counted 2.1% of total components.

Hierarchical clustering analysis. Hierarchical clustering of gene expression data is a more intuitive way to analyze the many different possible combinations of differentially expressed genes. **Fig. 4A** shows a two way clustering diagram of five groups of the triplicate samples based on the intensities of the quality filtered 30,849 probe sets on the arrays. The data showed again that the relationship within each group was closer than the relationship between different groups. Within each group, undifferentiated HepaRG cells had the least variation and liver tissues had the biggest variations. Between different groups, differentiated and undifferentiated HepaRG cells are more closely related to primary human hepatocytes than human liver tissues. HepG2 cells have the farther clustering distances to all other groups.

We further selected 115 genes annotated as being involved in xenobiotic metabolism, including the genes encoding Phase I and Phase II metabolizing enzymes and membrane transporters (a gene list and average signal intensities in log₂ scale from the three replicate sets are provided in Supplemental Table S3). The average signal intensities on the probes annotated to the selected genes were clustered in **Fig. 4B** shown in the groups of Phase I enzymes (ADHs, ALDHs, CYPs, and FMOs), Phase II enzymes (GSTs, NATs, SULTs, and UGTs), and

DMD #31831

transporters (ABCBs, ABCCs, ABCGs, and SLCOs). A similar clustering pattern as in the whole genome analysis was observed in the Phase-I and Phase-II metabolizing enzymes, indicated that HepG2 cells are the most dissimilar to the rest of the groups. A comparison of the expression values of these drug response genes among HepG2 cells, differentiated HepaRG cells, and primary human hepatocytes is also shown in Supplemental Figure S1 with means and standard deviations. Differences of the expression levels between HepG2 cells and primary human hepatocytes as well as between differentiated HepaRG cells and primary human hepatocytes were determined in a student *t* test. In overall, difference of gene expression in these drug response genes between HepG2 and primary human hepatocytes is much bigger than between HepaRG and primary human hepatocytes. For examples, all drug metabolizing cytochrome P450 genes are expressed at a significantly different level of *** $p < 0.001$, except CYP3A43 at ** $p < 0.01$, in HepG2 than in primary human hepatocytes (Figure S1-A), but only CYP2D6 is expressed at a significantly different level of *** $p < 0.001$ in HepaRG than in primary human hepatocytes and CYP1A2, CYP2A6, and CYP2C8 are at a significantly different level of ** $p < 0.01$. A similar situation is also found for many other drug response genes, such as ADH1A, ADH1B, ADH1C, ADH4, ALDH1L1, ALDH1L2, ALDH9A1, NAT1, NAT2, GSTA1, GSTA3, GSTK1, SULT1A1, SULT1A2, SULT2A1, UGT1A1, UGT1A6, UGT2B4, UGT2B15, UGT2B17, UGT3A1, ABCB1, ABCB4, ABCC10, SLCO2B1, etc (Figure S1).

HepaRG cells were derived from a human hepatocarcinoma liver tissue. Abnormality of the karyotype in HepaRG cells has been identified with a trisomic chromosome 7 and a translocated chromosome from 22 to 12 (Gripon et al., 2002). Here, we examined whether the karyotype abnormality has any influence on gene expression. Gene expression profiles across each chromosome were compared between differentiated HepaRG cells and primary human

DMD #31831

hepatocytes. Among the all chromosomes, only chromosome 7 had a significant higher gene expression level ($p < 0.001$ in a t -test) in differentiated HepaRG cells compared to primary human hepatocytes (**Fig. 5A**). The rest of the chromosomes, including the translocated chromosome 22 (**Fig. 5B**, $p = 0.084$ in a t -test), did not show significant difference of gene expression between differentiated HepaRG cells and primary human hepatocytes.

DMD #31831

Discussion

The current study used Affymetrix gene expression arrays to establish genome-wide gene expression profiles of HepaRG cells at both undifferentiated and differentiated stages and compared the genome-wide gene expression profiles of HepaRG cells and HepG2 cells with human primary hepatocytes and human liver tissues using similarity matrix, principal components, and hierarchical clustering methods. The comparison was also done for many drug response genes. These analyses conclude that the mRNA content in HepaRG cells more accurately reflects primary human hepatocytes and human liver tissues than HepG2 cells.

The similarity matrix analysis shows the relative high r values (0.949-0.996) between any two replicates within a same type of the samples. These high r values indicate that the gene expression profiles generated by the Affymetrix gene expression arrays are highly reproducible. It is not surprising that the r values between the replicate samples in the cultured HepG2 and HepaRG cells were higher than in primary human hepatocytes and human liver samples, because the cultured cells were consisted of a highly homogenous cell population with little environment-mediated perturbations, but primary human hepatocytes and human liver samples came from different individuals in which their gene expression could be influenced by many factors which can not be controlled in the experiment. Particularly, the r values within the liver tissue samples (0.949-0.971) are relatively lower than the r values within other groups, indicating that a certain degree of variations exists among the individual liver samples, which may be caused by interindividual variations or mRNA quality of the liver tissues.

When similarity is compared between different groups of the samples, differentiated HepaRG and undifferentiated HepaRG cells are still highly similar ($r = 0.966$ in **Fig. 1B**) with only a small proportion of the probes (~10% in **Fig. 2**) expressed differentially by more than 2 fold. A

DMD #31831

list of differentially expressed genes and pathways can be found in Supplemental Table S2. Drug processing genes are the most significantly up-regulated genes during HepaRG differentiation, including many Phase-I enzymes (such as CYP2C9, CYP2C19, CYP2E1, and CYP3A4) and Phase-II enzymes (for example, UDP-glucosyltransferases and glucosyltransferase). Although both undifferentiated and differentiated HepaRG cells are very similar to primary human hepatocytes at whole genome gene expression, differentiated HepaRG cells express xenobiotic processing genes more similar to primary human hepatocytes than undifferentiated HepaRG cells.

Similarity is also high between primary human hepatocytes and liver tissues ($r=0.920$ in Fig. 1B), but with about 20% of probes expressed differentially (**Fig. 2**). Although hepatocytes are the major types of cells in liver, making up to 70-80% of the mass of the liver, liver also consists of several other types of cells, such as cholangiocytes, endothelial cells, hepatic stellate cells, and kupffer cells, which have different gene expression profiles than hepatocytes. Other factors which can influence the measurement of gene expression in human livers are the procedures for harvest, treatment, and storage of liver tissue samples. Therefore, freshly isolated primary human hepatocytes should be considered as the key reference for comparison of gene expression between the *in vitro* cultured liver cells and the *in vivo* liver cells.

The similarity levels represented by the Pearson product-moment correlation coefficient (r) value and number of the non-differentially expressed probes were higher between HepaRG cells and primary human hepatocytes or liver tissues than between HepG2 cells and primary human hepatocytes or liver tissues, while the differences were lower between HepaRG cells and primary human hepatocytes or liver tissues than between HepG2 cells and primary human hepatocytes or

DMD #31831

liver tissues. These data indicated that HepaRG cells expressed genes at a genome-wide level more similar to primary human hepatocytes and human livers than HepG2 cells.

The above conclusion is also supported by principal component analysis, which confirms that the variations in gene expression at a whole genome level are contributed mainly from HepG2 cells compared to HepaRG cells, primary human hepatocytes, and liver tissues.

Hierarchical clustering analysis also shows that the association of gene expression at genome levels is closer in each type of the groups than between different types of the groups (**Fig. 4A**). Within each type of the groups, undifferentiated HepaRG cells have the closest association, whereas, liver tissue samples have the least association. Between different groups, undifferentiated HepaRG and differentiated HepaRG cells are close each other, and primary human hepatocytes and liver tissues are close each other. Then, HepaRG cells are closer to primary human hepatocytes and liver tissues than HepG2 cells. When a set of genes involved in drug response, including many Phase-I enzymes, Phase-II enzymes, and transporters were selected for a clustering analysis (**Fig. 4B**), the gene expression profiles are also more similar between HepaRG cells and human primary hepatocytes or liver tissues than between HepG2 cells and human primary hepatocytes or liver tissues. These findings are in agreement with several previous studies (Aninat et al., 2006; Le Vee et al., 2006; Richert et al., 2006; Kanebratt and Andersson, 2008a; Kanebratt and Andersson, 2008b). When the differences in expression levels among HepG2, HepaRG, and primary human hepatocytes were compared in the major drug response gene families, such as CYPs, ADHs, ALDHs, FMOs, NATs, GSTs, SULTs, UGTs, ABCBs, ABCCs, ABCGs, and SLCOs (Supplemental Figure S1), the differences between HepG2 and primary human hepatocytes are much larger than the differences between HepaRG and primary human hepatocytes for the most genes.

DMD #31831

In conclusion, we used a high-throughput genome-wide approach to define gene transcriptional profiles of HepaRG cells at both differentiated and undifferentiated stages and compared the gene expression profiles of HepaRG cells and HepG2 cells with human primary hepatocytes and liver tissues. We found gene transcription levels in HepaRG cells have a much higher level of similarity to human primary hepatocytes and liver tissues in comparison to HepG2 cells; the most commonly used cultured cells for studying liver biology. The transcriptomic similarity of HepaRG with human primary hepatocytes is encouraging for use of the HepaRG cells in the study of xenobiotic metabolism, hepatotoxicology, and hepatocyte differentiation in future. These sets of data can also serve as a database for researchers who want to compare expression levels of any genes in HepaRG cells, HepG2 cells, primary human hepatocytes, and human liver tissues.

The current study highlights the similarity of gene transcription between HepaRG cells and human primary hepatocytes or liver tissues in comparison with HepG2 cells. The high similarity at mRNA levels between HepaRG cells and human primary hepatocytes or liver tissues does not necessarily mean that the similarity occurs also at protein levels. Studies of genome-wide protein levels require high-throughput protein arrays which are not available yet.

It is also worth noting that because a trisomic chromosome 7 exists in HepaRG cells, genes located on chromosome 7 may have higher expression levels in HepaRG cells than in primary human hepatocytes due to the extra copy of chromosome 7. This factor should be taken into consideration when an experimental design in the use of HepaRG cells involves genes located on chromosome 7.

DMD #31831

Acknowledgements

We thank the Kansas University Medical Center-Microarray Facility (KUMC-MF) for generating array data sets. The Microarray Facility is supported by the Kansas University-School of Medicine, KUMC Biotechnology Support Facility, the Smith Intellectual and Developmental Disabilities Research Center (HD02528), and the Kansas IDeA Network of Biomedical Research Excellence (RR016475).

DMD #31831

References

- Aninat C, Piton A, Glaise D, Le Charpentier T, Langouet S, Morel F, Guguen-Guillouzo C and Guillouzo A (2006) Expression of cytochromes P450, conjugating enzymes and nuclear receptors in human hepatoma HepaRG cells. *Drug Metab Dispos* **34**:75-83.
- Glynn D, Sherman BT, Hosack DA, Yang J, Baseler MW, Lane HC, and Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology* **4**: P3.
- Gripon P, Rumin S, Urban S, Le Seyec J, Glaise D, Cannie I, Guyomard C, Lucas J, Trepo C and Guguen-Guillouzo C (2002) Infection of a human hepatoma cell line by hepatitis B virus. *Proc Natl Acad Sci U S A* **99**:15655-15660.
- Guillouzo A, Corlu A, Aninat C, Glaise D, Morel F and Guguen-Guillouzo C (2007) The human hepatoma HepaRG cells: a highly differentiated model for studies of liver metabolism and toxicity of xenobiotics. *Chem Biol Interact* **168**:66-73.
- Hewitt NJ, Lechon M, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Guillouzo A, Tuschl G, Li AP, LeCluyse E, Groothuis GMM and Hengstler JG (2007) Primary Hepatocytes: Current Understanding of the Regulation of Metabolic Enzymes and Transporter Proteins, and Pharmaceutical Practice for the Use of Hepatocytes in Metabolism, Enzyme Induction, Transporter, Clearance, and Hepatotoxicity Studies. *Drug Metabolism Reviews* **39**:159-234.
- Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B and Speed TP (2003) Summaries of Affymetrix GeneChip probe level data. *Nucl. Acids Res.* **31**:e15-.
- Kanebratt KP and Andersson TB (2008a) Evaluation of HepaRG cells as an in vitro model for human drug metabolism studies. *Drug Metab Dispos* **36**:1444-1452.

DMD #31831

Kanebratt KP and Andersson TB (2008b) HepaRG cells as an in vitro model for evaluation of cytochrome P450 induction in humans. *Drug Metab Dispos* **36**:137-145.

Kato M, Chiba K, Horikawa M and Sugiyama Y (2005) The quantitative prediction of in vivo enzyme-induction caused by drug exposure from in vitro information on human hepatocytes. *Drug Metab Pharmacokinet* **20**:236-243.

Lambert CB, Spire C, Claude N and Guillouzo A (2009a) Dose- and time-dependent effects of phenobarbital on gene expression profiling in human hepatoma HepaRG cells. *Toxicology and Applied Pharmacology* **234**:345-360.

Lambert CB, Spire C, Renaud M-P, Claude N and Guillouzo A (2009b) Reproducible chemical-induced changes in gene expression profiles in human hepatoma HepaRG cells under various experimental conditions. *Toxicology in Vitro* **23**:466-475.

Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe SM and Bor DH (2002) Timing of new black box warnings and withdrawals for prescription medications. *JAMA* **287**:2215-2220.

Le Vee M, Jigorel E, Glaise D, Gripon P, Guguen-Guillouzo C and Fardel O (2006) Functional expression of sinusoidal and canalicular hepatic drug transporters in the differentiated human hepatoma HepaRG cell line. *Eur J Pharm Sci* **28**:109-117.

LeCluyse E, Ajay M, Geraldine H, Kathy C, Ryan D and Andrew P (2000) Expression and regulation of cytochrome P450 enzymes in primary cultures of human hepatocytes. *Journal of Biochemical and Molecular Toxicology* **14**:177-188.

LeCluyse EL (2001) Human hepatocyte culture systems for the in vitro evaluation of cytochrome P450 expression and regulation. *European Journal of Pharmaceutical Sciences* **13**:343-368.

DMD #31831

Luo G, Guenthner T, Gan LS and Humphreys WG (2004) CYP3A4 induction by xenobiotics: biochemistry, experimental methods and impact on drug discovery and development. *Curr Drug Metab* **5**:483-505.

Ogino M, Nagata K and Yamazoe Y (2002) Selective suppressions of human CYP3A forms, CYP3A5 and CYP3A7, by troglitazone in HepG2 cells. *Drug Metab Pharmacokinet* **17**:42-46.

Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Van Deun K, Smith P, Berger B and Heller A (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* **32**:56-67.

Richert L, Liguori MJ, Abadie C, Heyd B, Manton G, Halkic N and Waring JF (2006) Gene expression in human hepatocytes in suspension after isolation is similar to the liver of origin, is not affected by hepatocyte cold storage and cryopreservation, but is strongly changed after hepatocyte plating. *Drug Metab Dispos* **34**:870-879.

Rodriguez-Antona C, Donato MT, Boobis A, Edwards RJ, Watts PS, Castell JV and Gomez-Lechon MJ (2002) Cytochrome P450 expression in human hepatocytes and hepatoma cell lines: molecular mechanisms that determine lower expression in cultured cells. *Xenobiotica* **32**:505-520.

Sassa S, Sugita O, Galbraith RA and Kappas A (1987) Drug metabolism by the human hepatoma cell, Hep G2. *Biochem Biophys Res Commun* **143**:52-57.

Wettenhall JM, Simpson KM, Satterley K and Smyth GK (2006) affyImGUI: a graphical user interface for linear modeling of single channel microarray data. *Bioinformatics* **22**:897-899.

DMD #31831

Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F and Spencer F (2004) A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association* **99**:909.

DMD #31831

Footnotes

This work was supported by National Institute of Health grant of 5P20 RR021940 and 1R01GM087376. Steven N. Hart is supported by a Madison and Lila Self Graduate Fellowship from the University of Kansas. Ye Li is supported by the University of Kansas Medical Center Biomedical Research Training Program.

*S. N. Hart and Y. Li contributed equally to this work.

DMD #31831

Figure Legends

FIG. 1. A. Similarity matrix of gene expression profiles for each pairwise comparison of HepG2 cells (HepG2-1, -2, -3), undifferentiated HepaRG cells (Undif HepaRG-1, -2, -3), differentiated HepaRG cells (Diff HepaRG-1, -2, -3), primary human hepatocytes (PHH-1, -2, -3), and human liver tissues (Liver-1, -2, -3). The number in each column represents Pearson's product-moment correlation coefficient r value. **B.** Average correlation coefficient r values for each type of the biological replicates within each group as well as between two groups. Data based on 30,849 probe sets passing a quality filtering test. The background colors in each column indicate different levels of the r values.

FIG. 2. Numbers and percentages of probe sets with differential gene expression by more than two fold between any two groups of the samples. The comparison was based on average signal intensities on each set of the probes from three replicates in each group of the samples.

FIG. 3. Principal components analysis on variations of gene transcription among HepG2 cells (HepG2), undifferentiated HepaRG cells (Undif HepaRG), differentiated HepaRG cells (Diff HepaRG), primary human hepatocytes (PHH), and liver tissues (Liver). For the 30,849 probes passing quality filtering, the relative contribution of the variance is shown by three major principal components plotted in three dimensions.

FIG. 4. A. Hierarchical clustering analysis of gene expression for HepG2 cells (HepG2-1, -2, -3), undifferentiated HepaRG cells (Undif HepaRG-1, -2, -3), differentiated HepaRG cells (Diff HepaRG-1, -2, -3), primary human hepatocytes (PHH-1, -2, -3), and human liver tissues (Liver-

DMD #31831

1, -2, -3). The clustering is based on the 30,849 probes passing quality filtering. **B.** Hierarchical clustering analysis of expression of Phase-I drug metabolizing enzyme genes (ADHs, ALDHs, CYPs, and FMOs), Phase-II drug metabolizing enzyme genes (GSTs, NATs, SULTs, and UGTs), and membrane transporter genes (ABCBs, ABCCs, ABCGs, and SLCOs). The clustering is based on average signal intensities from the three replicates in each group of the samples.

FIG. 5. Comparison of gene expression profiles across chromosome 7 (**A**) and 22 (**B**) between primary human hepatocytes (PHH) and differentiated HepaRG cells (Diff HepaRG). In the first two panels, the Y axis represents the signal intensity levels in \log_2 scale and the X axis indicates the genomic positions along the chromosomes where the microarray probes are located for PHH and Diff HepaRG, respectively. The third panel was constructed by subtracting PHH values from Diff HepaRG signal values for each probe on the chromosome and then clustered as in the first two panels. For example, if the expression of a given probe was 5 in PHH and 3 in Diff HepaRG, then that value would be $5-3=2$. Negative values indicate probes with lower expression in PHH than Diff HepaRG. Blue areas represent few data points, whereas red areas indicate more data points. In the forth panel, signal intensity levels from all probes on each chromosome are compared between Diff HepaRG and PHH with means and standard deviations. The differences are tested by a *t*-test. *** $p<0.001$ indicates that a significant difference is identified between the two sets of samples.

Fig. 1

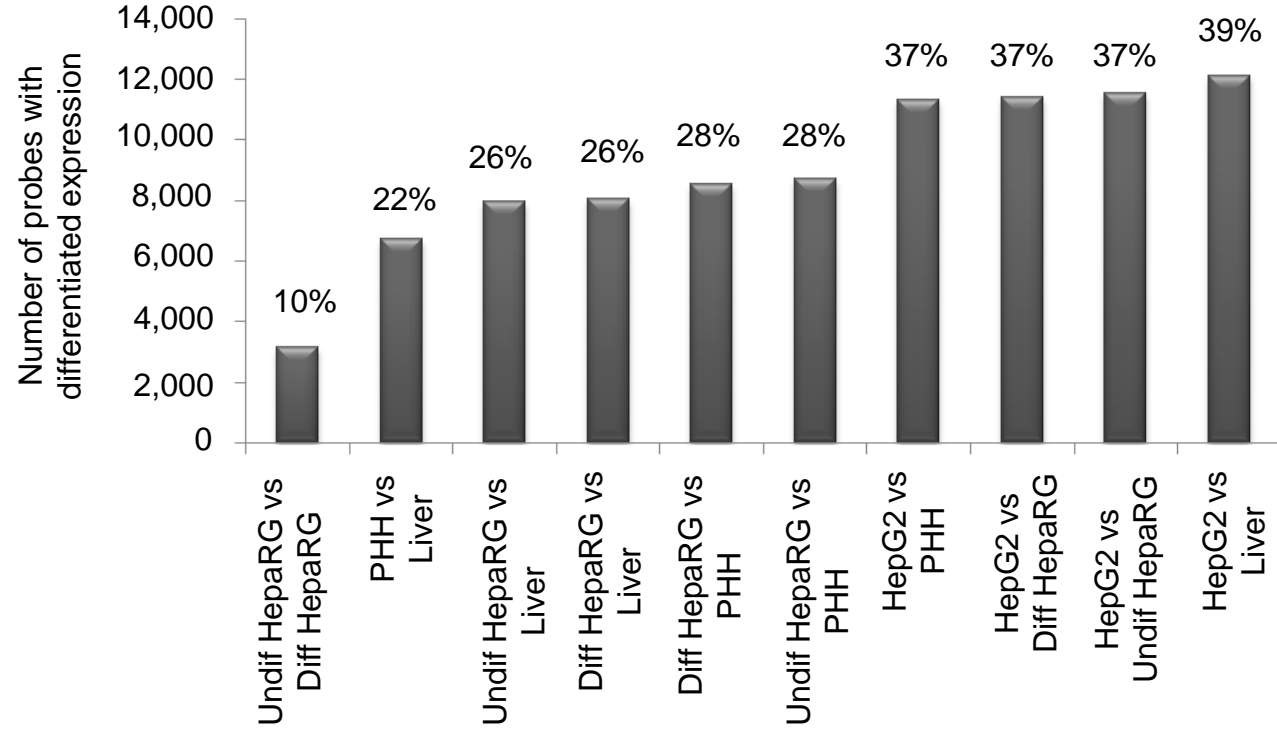
A

	HepG2-1	HepG2-2	HepG2-3	Undif HepaRG-1	Undif HepaRG-2	Undif HepaRG-3	Diff HepaRG-1	Diff HepaRG-2	Diff HepaRG-3	PHH-1	PHH-2	PHH-3	Liver-1	Liver-2	Liver-3
HepG2-1	-														
HepG2-2	0.995	-													
HepG2-3	0.991	0.991	-												
Undif HepaRG-1	0.811	0.816	0.807	-											
Undif HepaRG-2	0.808	0.813	0.802	0.995	-										
Undif HepaRG-3	0.810	0.815	0.805	0.995	0.996	-									
Diff HepaRG-1	0.818	0.823	0.812	0.965	0.964	0.965	-								
Diff HepaRG-2	0.817	0.821	0.811	0.965	0.966	0.966	0.995	-							
Diff HepaRG-3	0.818	0.823	0.813	0.967	0.969	0.968	0.989	0.990	-						
PHH-1	0.815	0.817	0.817	0.893	0.887	0.890	0.899	0.899	0.884	-					
PHH-2	0.811	0.814	0.812	0.893	0.887	0.890	0.900	0.900	0.885	0.979	-				
PHH-3	0.808	0.810	0.811	0.884	0.877	0.881	0.891	0.890	0.874	0.984	0.983	-			
Liver-1	0.785	0.789	0.780	0.887	0.884	0.886	0.886	0.888	0.879	0.928	0.937	0.919	-		
Liver-2	0.814	0.818	0.809	0.892	0.888	0.890	0.886	0.886	0.881	0.925	0.922	0.914	0.962	-	
Liver-3	0.775	0.780	0.768	0.873	0.871	0.873	0.875	0.878	0.867	0.908	0.922	0.901	0.971	0.949	-

B

	HepG2	Undif HepaRG	Diff HepaRG	PHH	Liver
HepG2	0.992				
Undif HepaRG	0.810	0.995			
Diff HepaRG	0.817	0.966	0.991		
PHH	0.813	0.887	0.891	0.982	
Liver	0.791	0.883	0.881	0.920	0.961

Fig. 2



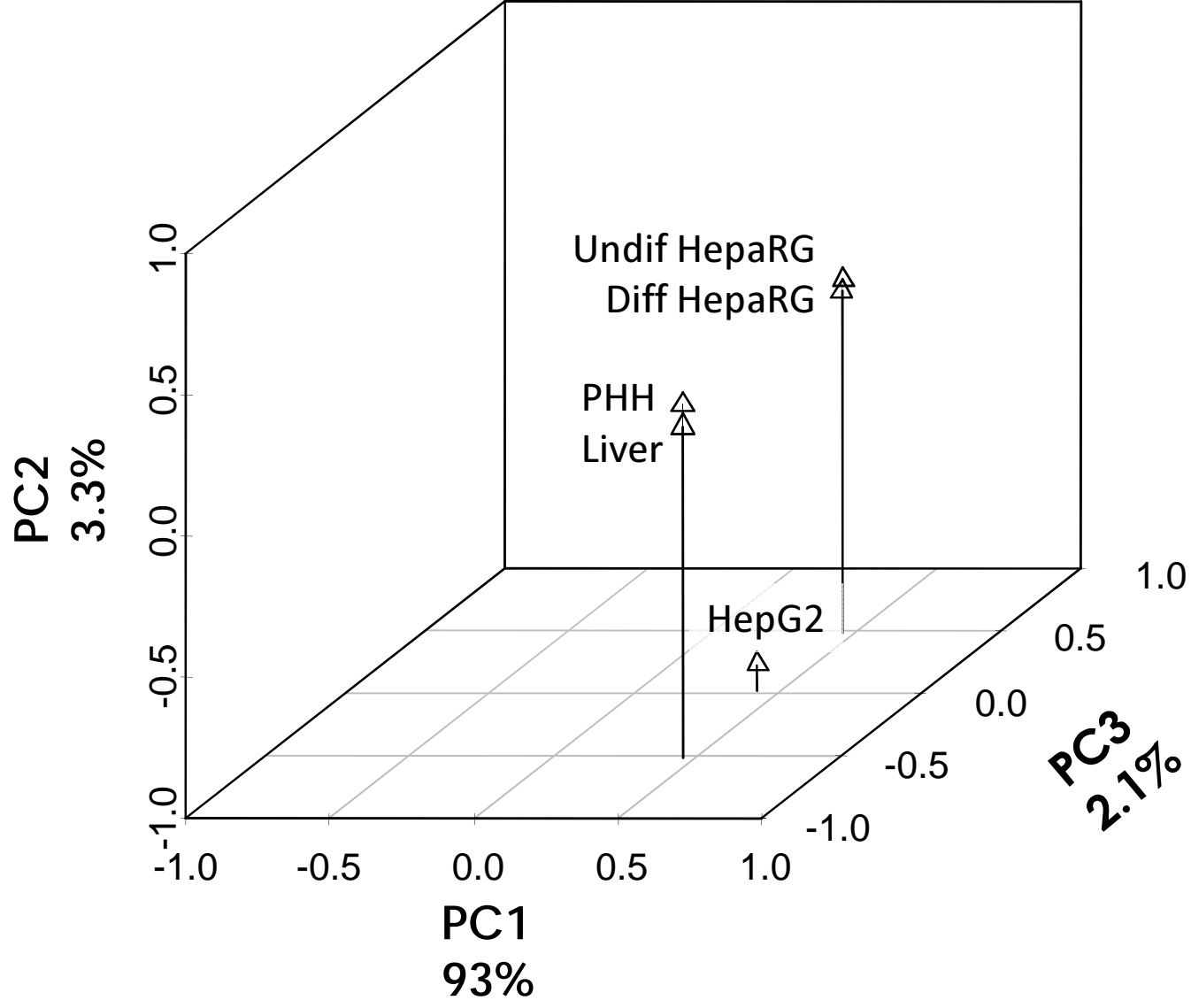


Figure 4

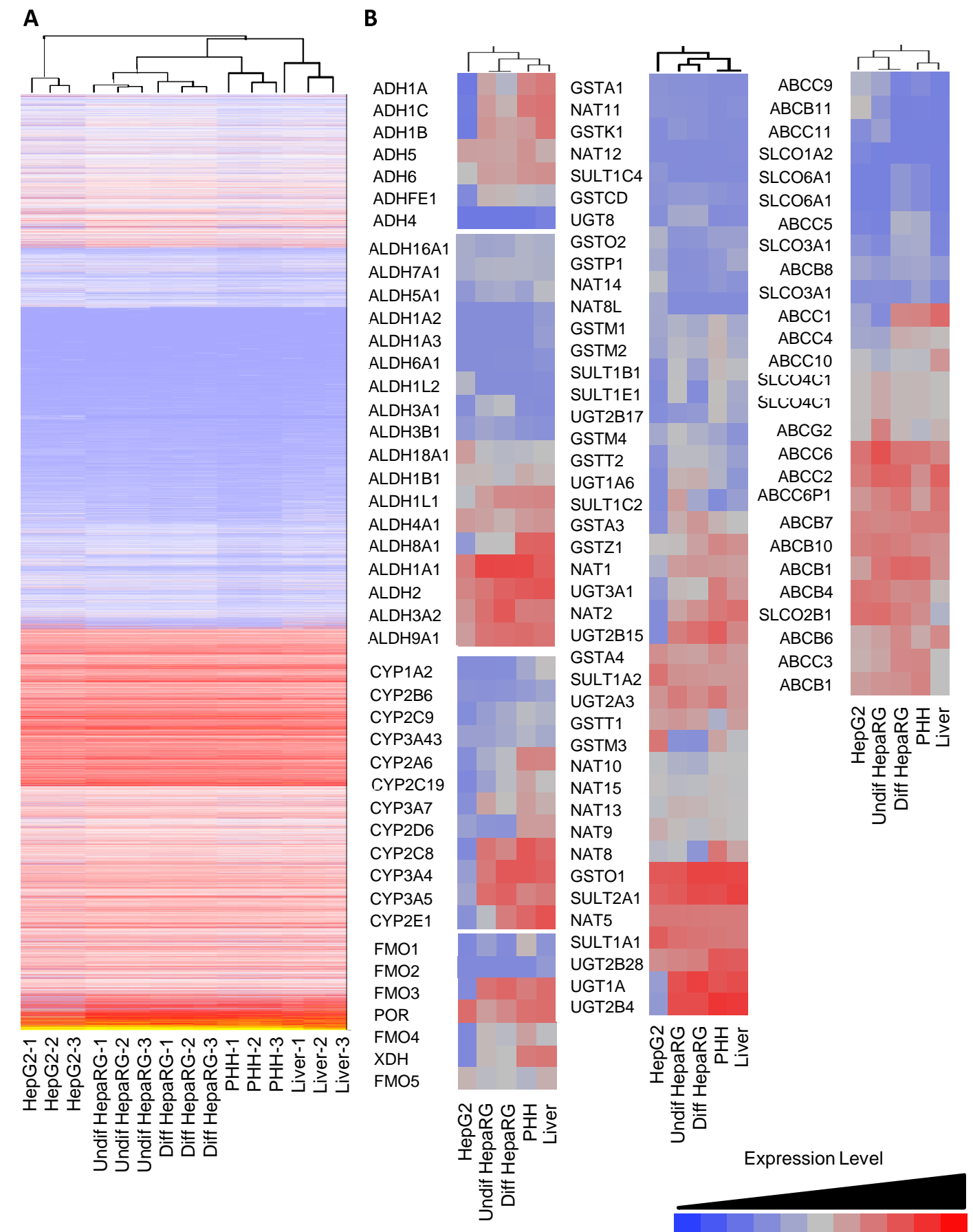


Fig. 5

