Qualifying a human Liver-Chip for predictive toxicology: Performance assessment and economic implications

Lorna Ewart[1*], Athanasia Apostolou[1], Skyler A. Briggs[1], Christopher V. Carman[1], Jake T. Chaff[1], Anthony R. Heng[1], Sushma Jadalannagari[1], Jeshina Janardhanan[1], Kyung-Jin Jang[1], Sannidhi R. Joshipura[1], Mahika Kadam[1], Marianne Kanellias[1], Ville J. Kujala[1], Gauri Kulkarni[1], Christopher Y. Le[1], Carolina Lucchesi[1], Dimitris V. Manatakis[1], Kairav K. Maniar[1], Meaghan E. Quinn[1], Joseph S. Ravan[1], Ann Catherine Rizos[1], John F.K. Sauld[1], Josiah Sliz[1], William Tien-Street[1], Dennis Ramos Trinidad[1], James Velez[1], Max Wendell[1], Onyi Irrechukwu[2], Prathap Kumar Mahalingaiah[3], Donald E. Ingber[4,5,6], Daniel Levner[1]

[1]Emulate Inc., 27 Drydock Avenue, Boston, MA, United States
[2]Janssen Pharmaceuticals, Spring House, PA, United States
[3]Investigative Toxicology and Pathology, Abbvie, North Chicago, IL, United States
[4]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, United States
[5]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, United States
[6]Vascular Biology Program and Department of Surgery, Harvard Medical School and Boston Children's Hospital, Boston, MA, United States

*Corresponding author: Lorna Ewart lorna.ewart@emulatebio.com

---

Abstract

Human organ-on-a-chip (Organ-Chip) technology has the potential to disrupt preclinical drug discovery and improve success in drug development pipelines as it can recapitulate organ-level pathophysiology and clinical responses. The Innovation and Quality (IQ) consortium, formed by multiple pharmaceutical and biotechnology companies, has published guidelines that define criteria for qualifying preclinical models, however, systematic and quantitative evaluation of the predictive value of Organ-Chips has not yet been reported. Here, 780 Liver-Chips were analyzed to determine their ability to predict drug-induced liver injury (DILI) caused by small molecules identified as benchmarks by the IQ consortium. The Liver-Chip met the qualification guidelines across a blinded set of 27 known hepatotoxic and non-toxic drugs with a sensitivity of 80% and a specificity of 100%. A computational economic value analysis suggests that with this performance the Liver-Chip could generate $3 billion annually for the pharmaceutical industry due to increased R&D productivity.

Introduction

Despite billion-dollar investments in research and development, the process of approving new drugs remains lengthy and costly due to high attrition rates[1,2,3]. Failure is common because the models used preclinically—which include computational, traditional cell culture, and animal models—have limited predictive validity[4]. The resulting damage to productivity in the pharmaceutical industry causes concern across a broad community of drug developers, investors, payers, regulators, and patients, the last of whom desperately need access to medicines with proven efficacy and improved safety profiles. Approximately 75% of the cost in research and development is the cost of failure[5]—that is, money spent on projects in which the candidate drug was deemed efficacious and safe by early testing but was later revealed to be ineffective or unsafe in human clinical trials. Pharmaceutical companies are addressing this challenge by learning from drugs that failed and devising frameworks to unite research and development organizations to enhance the probability of clinical success[6,7,8,9]. One of the major goals of this effort is to develop preclinical models that could enable a "fail early, fail fast" approach, which would result in candidate drugs with greater probability of clinical success, improved patient safety, lower cost, and a faster time to market.

There are significant practical challenges in ascertaining the predictive validity of new preclinical models, as there is a broad diversity of chemistries and mechanisms of action or toxicity to consider, as well as significant time needed to confirm the model's predictions once tested in the clinic. Consequently, arguments for the adoption of these new models are often based on features that are presumed to correlate with human responses to pharmacological interventions—realistic histology, similar genetics, or the use of patient-derived tissues. But even here there is a common problem in much of the academic literature: the important model features are chosen post-hoc by the authors and not prospectively by an independent third party that has expertise in the therapeutic problem at hand[10].

The Innovation and Quality (IQ) consortium is a collaboration of pharmaceutical and biotechnology companies that aim to advance science and technology to enhance drug discovery programs. To further this goal, the consortium has described a series of performance criteria that a new preclinical model must meet to become qualified. Within this consortium is an affiliate dedicated to microphysiological systems (MPS), which include organ-on-a-chip (Organ-Chip) technology that employs microfluidic engineering to recapitulate *in vivo* cell and tissue microenvironments in an organ-specific context[11,12]. This is achieved by recreating tissue-tissue interfaces and providing fine control over fluid flow and mechanical forces[13,14], optionally including supporting interactions with immune cells[15] and microbiome[16], and reproducing clinical drug exposure profiles[17]. Recognizing the promise of MPS for drug research and development, the IQ MPS affiliate has provided guidelines for qualifying new models for specific contexts of use to help advance regulatory acceptance and broader industrial adoption[18]; however, to this date, there have been no publications describing studies that carry out this type of performance validation for any specific context of use or that demonstrate an MPS capable of meeting these IQ consortium performance goals.

Guided by the IQ MPS affiliate's roadmap on liver MPS[19], which states that *in vitro* models for predicting drug-induced liver injury (DILI) that meet its guidelines are more likely to exhibit higher predictive validity than those that do not, we rigorously assessed commercially available human Liver-Chips (from Emulate, Inc.) within the context of use of DILI prediction. In this study, we tested 780 Liver-Chips using a blinded set of 27 different drugs with known hepatotoxic or non-toxic behavior recommended by the IQ consortium (Table 1). We compared the results to the historical performance of animal models as well as 3D spheroid cultures of primary human hepatocytes, which are preclinical models that are frequently employed in this context of use in the pharmaceutical industry[20]. In addition, we analyzed the Liver-Chip results from an economic perspective by estimating the financial value they could offer through their use in preclinical development in supporting toxicity-related decisions. We conclude with recommendations on how this type of platform might be implemented in pharmaceutical industry screening programs.

Drug	IQ MPS List	Tested in Spheroid	Spheroid False Negative	Garside DILI Rank
Ambrisentan	Yes, matched with Sitaxsentan	Yes	No	5
Asunaprevir	Yes, no matched pair	No	No	2
Benoxaprofen	No	Yes	Yes	1
Beta-Estradiol	No	Yes	Yes	3
Buspirone	Yes, matched with Nefazodone	Yes	No	4
Chlorpheniramine	No	Yes	Yes	3
Clozapine	Yes, matched with Olanzapine	Yes	No	2
Diclofenac	Yes, no matched pair	Yes	No	2
Entacapone	Yes, matched with Tolcapone	Yes	No	4
Fialuridine	Yes, matched with FIRU	Yes	No	1
FIRU	Yes, matched with Fialuridine	No	No	5
Labetalol	No	Yes	Yes	1
Levofloxacin	Yes, matched with Trovafloxacin	Yes	Yes	2
Lomitapide	No, Mipomersen substitute	No	No	3
Nefazodone	Yes, matched with Buspirone	Yes	No	1
Olanzapine	Yes, matched with Clozapine	No	No	5
Pioglitazone	Yes, matched with Troglitazone	Yes	Yes	3
Simvastatin	No	Yes	Yes	2
Sitaxsentan	Yes, matched with Ambrisentan	Yes	No	1
Stavudine	No	Yes	Yes	1
Tacrine	No	Yes	Yes	2
Telithromycin	Yes, no matched pair	No	No	1
Tolcapone	Yes, matched with Entacapone	Yes	No	1
Troglitazone	Yes, matched with Pioglitazone	Yes	No	1
Trovafloxacin	Yes, matched with Levofloxacin	Yes	No	1
Ximelagatran	No	Yes	Yes	1
Zileuton	Yes, no matched pair	Yes	Yes	2

Table 1. Small Molecule Drug Compounds used in the Liver-Chip evaluation. The 27 small molecule drugs are listed according to the IQ MPS affiliate classification and their ranking in the Garside DILI severity category, where 1 corresponds to drugs with severe clinical DILI and 5 to those with no DILI[26,42]. Structurally related toxic and non-toxic pairs are indicated too.

Results

Liver-Chip satisfies IQ MPS affiliate guidelines

The IQ guidelines for assessment of an *in vitro* liver MPS within the DILI prediction context of use requires evidence that the model replicates key histological structures and functions of the liver; furthermore, the model must be able to distinguish between seven pairs of small molecule toxic drugs and their non-toxic structural analogs. If the model passes through these hurdles, it must demonstrate its ability to predict the clinical responses of six additional selected drugs.

The Liver-Chips that we evaluated against these standards contain two parallel microfluidic channels separated by a porous membrane. Following the manufacturer's instructions, primary human hepatocytes are cultured between two layers of extracellular matrix (ECM) in the upper 'parenchymal' channel, while primary human liver sinusoidal endothelial cells (LSECs), Kupffer cells, and stellate cells are placed in the lower 'vascular' channel in ratios that approximate those observed *in vivo* (Figure 1a). All cells passed quality control criteria that included post-thaw viability >90%, low passage number (preferably P3 or less), and expression of cell-specific markers. Similar results were obtained using hepatocytes from two different human donors, which were procured from the same commercial vendor (Supplementary Table S1).

Live microscopy of the Liver-Chips revealed a continuous monolayer of hepatocytes displaying cuboidal and binucleated morphology in the upper 'parenchymal' channel of the chips, as well as a monolayer of polygonal shaped LSECs in the bottom 'vascular' channel, on the opposite side of the porous membrane (Figure 1b). Confocal fluorescence microscopy also confirmed liver-specific morphological structures as indicated by the presence of differentiation markers, including bile canaliculi containing a polarized distribution of F-actin and multidrug resistance-associated protein 2 (MRP2), hepatocytes rich with mitochondrial membrane ATP synthase beta subunit (ATPB), PECAM-1 (CD31) expressing LSECs, $CD68^+$ Kupffer cells, and desmin-containing stellate cells (Figure 1c). In addition, transmission electron microscopy confirmed the existence of similar cell-cell relationships and structures to those found in human liver, including well developed junction-lined bile canaliculi and adhesions between Kupffer cells and sinusoidal endothelial cells (Supplementary Figure S1).

Albumin and urea production are widely accepted as functional markers for cultured hepatocytes with the goal of reaching production levels observed in human liver *in vivo* (~ 20-105 µg and 56-159 µg per 10^6 hepatocytes per day, respectively)[19,21]. Liver-Chips fabricated with cells from two different hepatocyte donors were able to maintain physiologically relevant levels of albumin and urea synthesis over one week in culture (Figure 1d). Importantly, in line with the IQ MPS guidelines, the coefficient of variation for the mean daily production rate of urea was always below 15% in both donors; however, it was higher (<45%) for albumin production. These data corroborate the reproducibility and robustness of the Liver-Chip across experiments and highlight variability across donors that is not unlike the variability observed in humans. In fact, it is important to be able to analyze and understand donor-to-donor variability when evaluating cell-based platforms for the prediction of clinical outcome[22] or when a drug moves into clinical studies.

Because hepatocytes maintained in conventional static cultures rapidly reduce transcription of relevant liver-specific genes[23], the IQ MPS guidelines require confirmation that the genes representing major Phase I and II metabolizing enzymes, as well as uptake and efflux drug transporters, are expressed and that their levels of expression are stable. When analyzed on days 3 and 5 of culture, we detected high levels of expression of key genes defined by the IQ MPS guidelines, including CYP3A4, CYP2B6, CYP2C9, BSEP, and MRP2, confirming that gene transcription remained active within the chip (Figure 1e). Gene expression levels were not statistically different between day 3 or 5, apart from CYP2B6, CYP2C9, CYP2D6, and BSEP, which were significantly higher on day 5 compared to day 3, illustrating no significant degradation of activity. Previously, the same Liver-Chip has been shown to exhibit Phase I and II functional activities that are comparable to

freshly isolated human hepatocytes and 3D hepatic spheroids[21,25] as well as superior activity relative to hepatocytes in a 2D sandwich-assay plate configuration[21].

As these data confirmed that the Liver-Chip meets the major structural characterization and basic functionality requirements stipulated by the IQ MPS guidelines, we then carried out studies to evaluate this human model as a tool for DILI prediction. IQ MPS identified seven small-molecule drugs that have been reported to produce DILI in clinical studies as well as their structural analogs that are inactive or exhibit lower activity and do not produce clinical DILI (Table 1). Past work in the MPS field has focused on technically accessible endpoints that can be easily measured but are unfortunately not clinically relevant or translatable (e.g., IC_{50} for reduction in total ATP content)[26,27]. Furthermore, although cytotoxicity measures are fundamental in the assessment of a drug's potential for hepatotoxicity *in vitro* [28,29], gene expression and various phenotypic changes can occur at much lower concentrations[30,31]. As the Liver-Chip enables multiple longitudinal measures of drug effects and use of multiple measures may provide further sensitivity and add value[32], we assessed drug toxicities by quantifying both inhibition of albumin production as a general measure of hepatocyte viability and increases in release of alanine aminotransferase (ALT) protein, which is used clinically as a measure of liver damage. We also scored hepatocyte injury using morphological analysis at 1, 3, and 7 days after drug or vehicle exposure, where higher injury scores indicate greater damage.

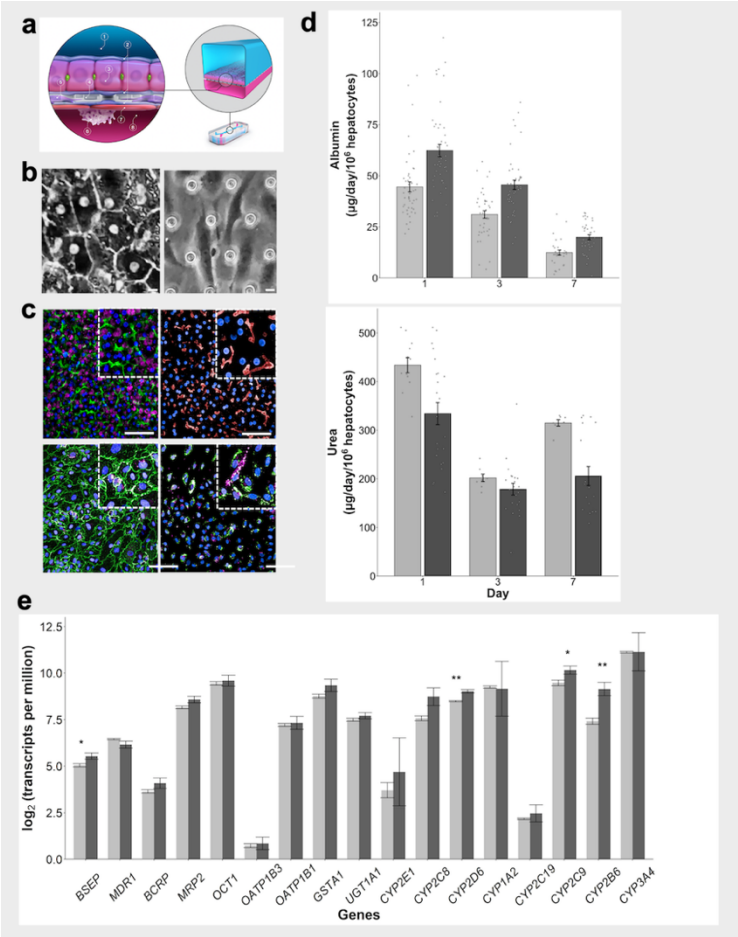

Figure 1 Recapitulation of human liver structure and function in the Liver-Chip. a) Schematic of the Liver-Chip showing primary human hepatocytes (3) that are sandwiched within an extracellular matrix (2) on a porous

membrane (4) within the upper parenchymal channel (1), while human liver sinusoidal endothelial cells (7), Kupffer cells (6), and stellate cells (5) are cultured on the opposite side of the membrane in the lower vascular channel (8). **b)** Representative phase contrast microscopic images of hepatocytes in the upper channel of Liver-Chip (left) and non-parenchymal cells in the lower vascular channel (right); the regular array of circles are the pores in the membrane in the right image. **c)** Representative immunofluorescence microscopic images showing the phalloidin stained actin cytoskeleton (green) and ATPB containing mitochondria (magenta) (top left), MRP2-containing bile canaliculi (red) (top right), CD31-stained liver sinusoidal endothelial cells (green) and desmin-containing stellate cells (magenta) (bottom left), and CD68+ Kupffer cells (green) co-localized with desmin-containing stellate cells (magenta). All images also show DAPI-stained nuclei (blue) (bar, 100 µm; inset is shown at 5 times higher magnification). **d)** Albumin (top) and urea (bottom) levels in the effluent from the upper channels of control Liver-Chips created with cells from 2 different donors (light and dark grey bars) on days 1, 3 and 7 measured by ELISA. Data are presented as mean $\pm$ standard deviation (S.D.) with N = 29 to 46 and N = 7 to 12 chips for albumin and urea respectively per donor. **e)** Levels of key liver-specific genes in control Liver-Chips as determined by RNA-seq analysis on days 3 (light grey) and 5 (dark grey). Data are presented as mean $\pm$ S.D. with N = 4 chips; statistical significance of values between day 3 and 5 was determined using an unpaired t-test; **, $p < 0.01$.

We tested the seven toxic drugs across eight concentrations that bracket the human plasma C_{max} for each drug based on free (non-protein bound) drug concentrations, with the highest concentrations at 300x C_{max} (unless not permitted by solubility limits) to represent clinically relevant test concentrations for *in vitro* models[33] (Supplementary Table S2). The known toxic compounds showed clear concentration- and time-dependent patterns that varied depending on compound. Typically, when albumin production was inhibited, morphological injury scores and ALT levels also increased, but we found that a decrease of albumin production was the most sensitive marker of hepatocyte toxicity in the Liver-Chip, as shown in sample paired comparisons of clozapine and olanzapine, troglitazone and pioglitazone, and trovafloxacin and levofloxacin (Figure 2a). Importantly, all seven of the toxic drugs reduced albumin production or resulted in an increase in ALT protein or injury morphology scores at lower multiples of the free human C_{max} compared to each of their non-toxic comparators (Table 2). Furthermore, immunofluorescence microscopic imaging for markers of apoptotic cell death (caspase 3/7) and mitochondrial injury measured by visualizing reduction of tetramethylrhodamine methyl ester (TMRM) accumulation (Figure 2b) provided confirmation of toxicity and, in many cases, provided some insight into the potential mechanism of toxicity. For example, the third-generation anti-infective trovafloxacin is believed to have an inflammatory component to its toxicity, potentially mediated by Kupffer cells, but this is only seen in animal models if an inflammatory stimulant such as lipopolysaccharide (LPS) is co-administered[34]. Interestingly, immunofluorescence microscopic imaging of the Liver-Chip revealed that there was a concentration-dependent increase in caspase 3/7 staining following trovafloxacin treatment (Figure 2b, bottom); this supports a potential apoptotic component to its toxicity. Interestingly, this was detected in the Liver-Chip without an inflammatory stimulant. Of note, Levofloxacin, the lesser toxic structural analog, did not cause cellular apoptosis.

Drug	Albumin IC_{50}		ALT		Morphology		IF imaging	
	Donor 1	Donor 2	Donor 1	Donor 2	Donor 1	Donor 2	Donor 1	Donor 2
Clozapine	34	87	300	> 300	100	100	Apoptosis	Apoptosis
Olanzapine	> 300	> 300	> 300	> 300	> 300	> 300	Apoptosis	No findings
Fialuridine	<0.1	<1	> 300	> 300	> 300	> 300	No findings	Steatosis
FIRU	> 300	> 300	> 300	> 300	> 300	> 300	No findings	Steatosis
Nefazodone	115	-	300	-	300	-	No findings	-
Buspirone	282	-	> 300	-	> 300	-	Apoptosis	-
Sitaxsentan	2.9	3.0	300	300	100	100	Mitotoxicity	Mitotoxicity
Ambrisentan	> 300	> 1000	> 300	> 1000	> 300	> 1000	No findings	No findings
Tolcapone	4.5	44	100	30	10	100	Mitotoxicity	Mitotoxicity
Entacapone	> 300	-	> 300	-	> 300	-	No findings	-
Troglitazone	10	125	300	300	300	300	Apoptosis	Apoptosis
Pioglitazone	268	> 300	> 300	> 300	> 300	> 300	No findings	No findings
Trovafloxacin	71	75	100	100	100	100	Apoptosis	Apoptosis
Levofloxacin	> 15	> 45	> 15	> 45	> 15	>45	No findings	No findings

Table 2. Data for the matched pair analysis proposed by IQ MPS guidelines. Data are presented as the lowest unbound human C_{max} multiplier causing a 50% reduction in albumin production, or the lowest concentration causing an increase in ALT protein or morphology score.

Together, these data support the Liver-Chip's value as a predictor of drug-induced toxicity in the human liver and demonstrate that this experimental system meets the basic IQ MPS criteria for preclinical model functionality. However, in addition to the seven matched pairs, the IQ MPS guidelines require that an effective human MPS DILI model can predict liver responses to six additional small-molecule drugs associated with clinical DILI. We only analyzed the effects of five of these drugs (diclofenac, asunaprevir, telithromycin, zileuton, and lomitapide) because the reported mechanism of toxicity of one of them (pemoline) is immune-mediated hypersensitivity[35], and this would potentially require a more complex configuration of the Liver-Chip containing additional immune cells. We also were unable to obtain one of the suggested drugs, mipomersen, from any commercial vendor; however, we tested lomitapide as an alternate, as both produce steatosis by altering triglyceride export, and lomitapide is known to induce elevated ALT levels[36].

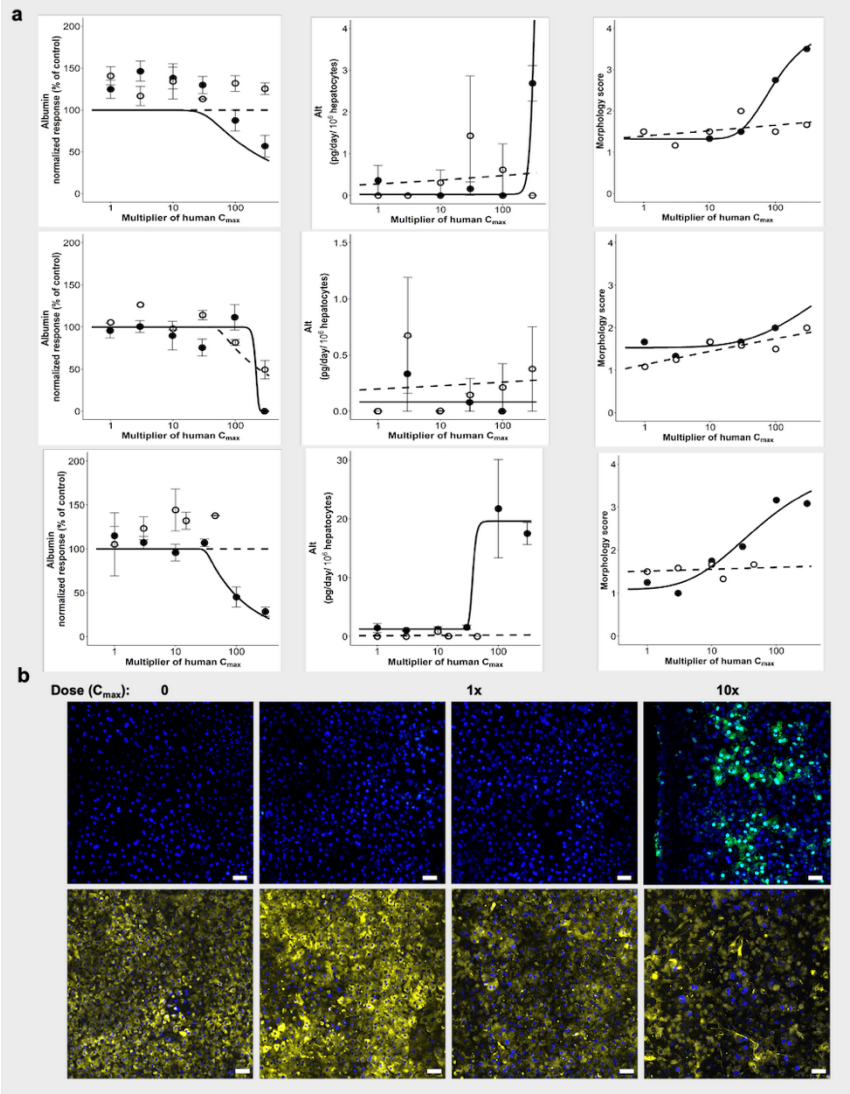

Figure 2. Detection of drug concentration-dependent toxicity and liver injury. a) Albumin (left), ALT (middle) and morphological injury score (right) for known non-toxic drugs (open circles) and their closely related toxic partner compounds (open circles) measured on day 3. Clozapine and olanzapine are shown at the top, troglitazone and pioglitazone below, and trovafloxacin and levofloxacin in the bottom row. **b)** Immunofluorescence microscopic images showing concentration-dependent increases in caspase 3/7 staining (green) indicative of apoptosis after treatment with trovafloxacin at 0, 1, 10, and 100 times the unbound human Cmax for 7 days (top). The bottom panel shows a concentration-dependent decrease in TMRM staining (yellow) indicative of mitotoxicity in response to treatment with sitaxsentan under similar conditions.

Results obtained with these drugs are presented in Table 3, with toxicity values indicating the lowest concentration at which toxicity was detected. Lomitapide was highly toxic when tested over all included concentrations down to 0.1x human plasma C_{max}, with all Liver-Chips showing signs of toxicity following five days of dosing. While telithromycin displayed a decrease in albumin along with a concomitant increase in ALT and morphological injury score, diclofenac and asunaprevir induced concentration and time-dependent changes in albumin and injury scores, but no elevation of ALT was seen with these drugs. Hepatotoxicity was also confirmed with immunofluorescence microscopy, which revealed apoptosis-mediated cell death following exposure to diclofenac, asunaprevir, or telithromycin. However, the Liver-Chip was unable to detect hepatotoxicity caused by zileuton, a

treatment intended for asthma. The exact mechanism of toxicity of zileuton is unknown, but it likely involves production of intermediate reactive metabolites due to oxidative metabolism by the cytochrome P450 isoenzymes 1A2, 2C9, and 3A4[37]. Although zileuton is >93% plasma protein bound[38], we do not believe this was responsible for the lack of toxicological effect, as we were able to detect toxicities induced by other highly protein-bound drugs in the test set.

Drug	Albumin IC_{50}		ALT		Morphology		IF imaging	
	Donor 1	Donor 2	Donor 1	Donor 2	Donor 1	Donor 2	Donor 1	Donor 2
Asunaprevir	190	-	> 1000	-	1000	-	Apoptosis	-
Benoxaprofen	8	20	> 100	100	< 0.1	100	No findings	Apoptosis
Beta-Estradiol	> 300	-	> 300	-	> 300	-	No findings	-
Chlorpheniramine	> 300	-	> 300	-	> 300	-	No findings	-
Diclofenac	384	-	> 1000	-	1000	-	Apoptosis	-
Labetalol	26	40	> 100	100	> 100	100	Apoptosis	Apoptosis
Lomitapide	4	-	> 1000	-	< 0.1	-	No findings	-
Simvastatin	< 0.1	<10	> 300	> 300	> 300	>300	No findings	Apoptosis
Stavudine	>60	132	300	> 300	300	300	No findings	No findings
Tacrine	76	> 25	> 300	> 1000	> 300	> 1000	Apoptosis	Apoptosis
Telithromycin	13	-	30	-	100	-	Apoptosis	-
Ximelagatran	32	> 300	30	> 300	100	300	No findings	No findings
Zileuton	> 100	210	> 100	> 300	> 100	> 300	No findings	No findings

Table 3. **Results obtained with the expanded drug list.** Data are presented as the lowest unbound human C_{max} multiplier causing a 50% reduction in albumin production, or the lowest concentration causing an increase in ALT release or morphology score.

Improved sensitivity for DILI prediction compared to spheroids and animal models

After fulfilling the major criteria of the IQ MPS affiliate guidelines, we considered the Liver-Chip to be qualified as a suitable tool to predict DILI during preclinical drug development. However, we wanted to also quantify the performance of the Liver-Chip in the predictive toxicology context. To do so, we expanded the drug test to include eight additional drugs (benoxaprofen, beta-estradiol, chlorpheniramine, labetalol, simvastatin, stavudine, tacrine, and ximelagatran) that were found to induce liver toxicity clinically, despite having gone through standard preclinical toxicology packages involving animal models prior to first-in-human administration. Importantly, the toxicities of these eight drugs have been shown to be poorly predicted by hepatic spheroids[26,27].

We proceeded to quantify any observed toxicity across the combined and blinded 27-drug set (Table 1) as a margin of safety (MOS)-like figure by taking the ratio of the minimum toxic concentration observed to the clinical C_{max}. We obtained the minimum toxic concentration by taking the lowest concentration identified by each of the primary endpoints—i.e., IC_{50} values for the decrease in albumin production, the lowest concentration at which we observed an increase in ALT protein, and the lowest concentration at which we observed injury via morphology scoring (Table 4). Minimum toxic concentrations generally corresponded to day 7 values, although day 3 values were occasionally lower. We then compared the MOS-like figures against a threshold value of 50 to categorize each compound as toxic or non-toxic, as previously reported for 3D hepatic spheroids, which used a similar threshold[26]. Analyzed in this manner, we found that, in addition to the drugs assessed as part of the IQ MPS-related analysis, the Liver-Chip correctly determined labetalol and benoxaprofen to be hepatotoxic, a response that was consistent across both donors and was indicated primarily by a reduction in albumin production. However, we found that simvastatin and ximelagatran were only toxic in one of the two donors, again showing the importance of including multiple donors during the risk assessment process. Overall, the Liver-Chip correctly predicted toxicity in 12 out of 15 toxic drugs that were tested using two donors, yielding a sensitivity of 80% on this drug set. This was almost double the sensitivity of 3D hepatic spheroids for the same drug set (42%)

based on previously published data[26,39,40], a preclinical model that is currently widely used in pharma and was only able to correctly identify eight out of the 19 toxic drugs in the set (Table 5a). Importantly, the Liver-Chip also did not falsely mark any drugs as toxic (specificity of 100%), whereas the 3D hepatic spheroids did (only 67% specificity)[26]; such false positives can significantly limit the usefulness of a predictive screening technology because of the profound consequences of erroneously failing safe and effective compounds. Interestingly, the three drugs not detected by Liver-Chip—levofloxacin, stavudine, and tacrine—were not detected as toxic drugs in spheroids either, suggesting that the Liver-Chip may subsume the sensitivity of spheroids and that their toxicities could involve other cells or tissues not present in these models. It is important to note that each of the toxic drugs tested was historically evaluated using animal models, and in each case the considerations and thresholds were deemed relevant for that drug to have an acceptable therapeutic window and thus progress into clinical trials. The ability of the Liver-Chip to flag 80% of these drugs for their DILI risk at their clinical concentrations represents a significant improvement in model sensitivity that could drive better decision making in preclinical development.

Drug	Chip MOS Donor 1	Chip MOS Donor 2	Chip MOS Both Donors	Spheroid MOS
Ambrisentan	> 3	> 10	> 10	> 127
Asunaprevir	4	-	-	-
Benoxaprofen	0.001	0.3	0.001	> 0.7
Beta-Estradiol	> 5	-	-	22500
Buspirone	> 15	-	-	16300
Chlorpheniramine	> 200	-	-	2141
Clozapine	1.8	5	1.8	14.5
Diclofenac	1.8	-	-	6.1
Entacapone	> 6	-	-	46.5
Fialuridine	0.1	2.8	0.1	12.3
FIRU	> 108	> 108	> 108	-
Labetalol	26	22	22	0.41
Levofloxacin	> 11.3	> 33.7	> 33.7	> 20
Lomitapide	0	-	-	-
Nefazodone	1.4	-	-	6.8
Olanzapine	> 20	> 2	> 20	-
Pioglitazone	2.8	> 2.8	2.8	> 5.3
Simvastatin	17	45	17	460
Sitaxsentan	0.04	1.6	0.04	8.7
Stavudine	247	107	107	> 144
Tacrine	> 12	> 12	> 12	696
Telithromycin	6	-	-	-
Tolcapone	0.004	0.05	0.004	0.3
Troglitazone	0.03	0.1	0.03	2.3
Trovafloxacin	20	19	19	> 24.9
Ximelagatran	30	300	30	335
Zileuton	> 7	15	15	> 7.7

Table 4. Calculation of Margin of Safety (MOS)-Like Figures. The analysis was carried out as described previously[26].

We examined each of the toxic drugs that were missed by the Liver-Chips to identify opportunities for future improvement. Using the threshold of 50 for determining toxicity, which we chose to compare our results to those from a past hepatic spheroid study, led to stavudine being classified a false negative. In fact, Liver-Chips do capture stavudine's toxicity at a higher threshold without introducing false positives, as described below. Tacrine is a reversible acetylcholinesterase inhibitor that undergoes glutathione conjugation by the phase II metabolizing enzyme glutathione S-transferase in liver. Polymorphisms in this enzyme can impact the amount of oxidative DNA damage, and the M1 and T1 genetic polymorphisms are associated with greater hepatotoxicity[41]. It is not known if either of the two hepatocyte donors used in this investigation have these polymorphisms, but the Liver-Chip was able to detect increased caspase 3/7 staining—indicative of apoptosis at the highest tested concentrations—although these changes were not associated with any release of ALT or decline in albumin. Levofloxacin, a fluoroquinolone antibiotic, was proposed by the IQ MPS affiliate as a lesser hepatotoxin compared to its structural analog trovafloxacin, but it is classified as a high clinical DILI concern in Garside's DILI severity category labeling[42]. Indeed, there are documented reports of hepatotoxicity with levofloxacin, but these occurred in individuals aged 65 years and above[43], and a post-market surveillance report documented the incidence of DILI to be less than one in a million people[44]. It is therefore reasonable to assume that the negative findings in both the Liver-Chip and spheroids may correctly represent clinical outcome.

Accuracy improved by accounting for drug-protein binding

When calculating the MOS-like values in the preceding section, we followed the published methods used for evaluating 3D hepatic spheroids[26], but these do not consider protein binding. Because the fundamental principles of drug action dictate that free (unbound) drug concentrations drive drug effects, we explored an alternative methodology for calculating the MOS-like values by accounting for protein binding using a previously reported approach[31]. Accordingly, we reanalyzed the findings for the 27 drugs in our study by accounting for protein binding. We compared the free fraction of drug concentration dosed in the Liver-Chip employing a medium containing 2% fetal bovine serum to the free fraction of the plasma C_{max}. By reanalyzing the Liver-Chip results using this approach and setting the threshold value to 375 (which we selected to maximize sensitivity while avoiding false positives), we obtained improved chip performance: a true positive rate (sensitivity) of 77% and 73% in donors one and two, respectively, and a true negative rate (specificity) of 100% in both donors (Table 5b). Importantly, the sensitivity increased to 87% when including the 18 drugs tested in both donors, and this enabled detection of stavudine's toxicity. Applying the same analysis to spheroids and similarly selecting a threshold to maximize sensitivity while maintaining 100% specificity yielded a sensitivity of only 47%. Remarkably, the Spearman correlation between the two-donor Liver-Chip assay and the Garside DILI severity scale yielded a value of 0.78 when using the protein binding-corrected analysis, whereas it was only 0.43 when using the lower threshold. Thus, the protein-binding-corrected approach not only produces higher sensitivity but also rank-orders the relative toxicity of drugs in a manner that corresponds better with the DILI severity observed in the clinic. This observation supports the validity of this analysis approach and its superiority over the uncorrected version. In short, these results provide further confidence that the Liver-Chip is a highly predictive DILI model and is superior in this capacity to other currently used approaches.

Model	True Positive	True Negative	False Positive	False Negative	Sensitivity (%)	Specificity (%)
Chip donor 1	16	5	0	6	73 [51.4-86.8%]	100
Chip donor 2	9	3	0	6	60 [35.3-80.3%]	100
Chip both donors	12	3	0	3	80 [54.2-92.8%]	100
Spheroid	8	2	1	11	42 [22.9-64.0%]	100

Model	True Positive	True Negative	False Positive	False Negative	Sensitivity (%)	Specificity (%)
Chip donor 1	17	5	0	5	77 [56.1-89.8%]	100
Chip donor 2	11	3	0	4	73 [47.5-89.0%]	100
Chip both donors	13	3	0	2	87 [61.5-96.0%]	100
Spheroid	9	3	0	10	47 [27.0-68.5%]	100

Table 5. Sensitivity and Specificity Determination (a) Predictive performance for chips and spheroids as determined by published analysis [26], which does not account for protein binding, and setting the threshold on the MOS-like values at 50 for both chips and spheroids (a value above 50 would indicate a drug as negative for toxicity) and (b) Predictive performance as determined by considering free (unbound) drug concentrations and setting the threshold on the MOS-like values at 375 for chips and at 2250 for spheroids. The 95% confidence intervals are shown for the sensitivity values.

The economic value of more predictive toxicity models in preclinical decision making

In addition to increasing patient safety, better prediction of drug toxicities can significantly improve the economics of drug development by driving a reduction in clinical trial attrition and promoting an increase in research and development productivity. To estimate the potential economic impact of the Liver-Chip considering its enhanced predictive validity, we sought external advice on financial analysis from Dr. Jack Scannell, who provided an economic value model of drug development as driven by decision quality during preclinical development (personal communication). We then parameterized the model using the results of the present study to compute the potential economic impact of the Liver-Chip. The financial model operates by tracking the costs of a simulated portfolio of clinical assets as it progresses through clinical trials. It follows Scannell & Bosley's[4] approach to replace conventional pipeline attrition parameters[5] with parameters that capture decision performance, true positive rates, false positive rates, and candidate quality. Using this method, and in contrast to conventional attrition-based pipeline models, we capture the consequences of earlier decisions in the development pipeline. The portfolio in the model starts as a representative mix of safe/unsafe and efficacious/non-efficacious drugs that are progressively filtered through each test phase, with the decision parameters set to recapitulate published attrition rates. Improvements in the predictive validity of preclinical safety testing are captured through their effects on the simulated portfolio entering the calculation: better preclinical safety testing leads to proportionally fewer unsafe drugs in the mix.

To estimate the economic impact of incorporating the Liver-Chip into preclinical research, we observed that DILI currently accounts for 13% of clinical trial failures that are due to safety concerns[46]. The present study revealed that the Liver-Chip, when used with two donors and analyzed with consideration for protein binding, provides sensitivity of 87% when applied to compounds that evaded traditional safety workflows. Combining these figures suggests that using the human Liver-Chip to test for DILI risk could lead to 10.4% fewer toxic drugs entering clinical trials. Remarkably, the model predicts that utilizing the Liver-Chip across all small-molecule drug development programs for this single context of use could generate the industry $3 billion annually due to increased research and development (R&D) productivity. If similar sensitivity is assumed for Organ-Chips that address four additional causes of safety failures—cardiovascular, neurological, immunological, and gastrointestinal toxicities, which together with DILI account for 80% of trial failures due to safety concerns—the model estimates that Organ-Chip technology could generate the industry over $24 billion annually if equally predictive as the Liver-Chip. These figures present a compelling economic incentive for the adoption of Organ-Chip technology alongside considerations of patient safety and the ethical concerns of animal testing.

Discussion

Organ-Chip technology has tremendous potential to revolutionize drug discovery and development[47], and many major pharmaceutical

companies have already invested in the technology, but routine utilization is limited[48]. This may be due to several factors, but the overriding fact is either that there has not been an end-to-end investigation showing that Organ-Chips replicate human biological responses in a robust and repeatable manner, that its performance exceeds that of existing preclinical models, or that there is a way to implement the technology into routine drug screening projects. This investigation directly addressed these three concerns. Furthermore, the broader stakeholder group—especially budget holders—need assurance that there will be a return on investment and that such technologies will help reverse the pharmaceutical industry's widely documented productivity crisis.

To counteract the R&D productivity crisis[49], the pharmaceutical industry is seeking physiologically relevant models that can be incorporated into the early-stage drug discovery programs in which the cost of attrition is lower and, ultimately, the quality of drug candidates progressing into the clinic will be higher[50,51]. Organ-on-a-chip technology utilizes microengineering to develop physiologically complex but human-relevant models; therefore, this technology should be implemented into programs to achieve this goal. To date, there has been no systematic evaluation of the validity of Organ-Chips or any other MPS for DILI prediction against criteria designed by a third party of experts. To our knowledge, no MPS has been evaluated against 27 small-molecule drugs in a single study involving two different human donors and hundreds of experiments, making this study the largest reported evaluation of Organ-Chip performance. The Liver-Chip has demonstrated that it can correctly distinguish toxic drugs from their non-toxic structural analogs, and, across a blinded set of 27 small molecules, has a true positive rate of 87%, a specificity of 100%, and a Spearman correlation of 0.78 against the Garside DILI severity scale when two donors are used, and data are corrected for protein binding. Importantly, these data were independently verified by two external toxicologists. Said differently, the Liver-Chip detected nearly seven out of every eight drugs that proved hepatoxic in clinical use despite having been deemed to have an appropriate therapeutic window by animal models; the Liver-Chip similarly detected two out of four such drugs that were additionally missed by 3D hepatic spheroids. Hence, we believe that these findings advocate the routine use of the human Liver-Chip in drug discovery programs to enhance the probability of clinical success while improving patient safety. This would be achieved by more-accurately categorizing risk associated with a candidate drug to provide valuable data to support a 'weight-of-evidence' argument both for entry into the clinic as well as for starting dose in phase I. Such added evidence could potentially remove any safety factor applied because of a liver finding in an animal model[52,53]. In turn, this would reduce overall cost and time in the preclinical development process.

A unique feature of this work is the demonstration of the throughput capability of Organ-Chip technology using automated culture instruments, as a total of 780 chips were created and analyzed. In terms of establishing effective workflows, scientists were placed into three teams: the first team prepared the drug solutions and supplied them in a blinded manner to the second team. The second team seeded, maintained, and dosed the Liver-Chips while carrying out various morphological, biochemical, and genetic analyses at the end of the experiment. The third team collected the effluents and performed real-time analyses of albumin and ALT as well as terminal immunofluorescence imaging using an automated confocal microscope (Opera Phenix; Perkin Elmer). In this manner, we were able to analyze and report the hepatotoxic effects of 27 drugs in 780 Liver-Chips that used cells from two human donors in a period of 16 weeks.

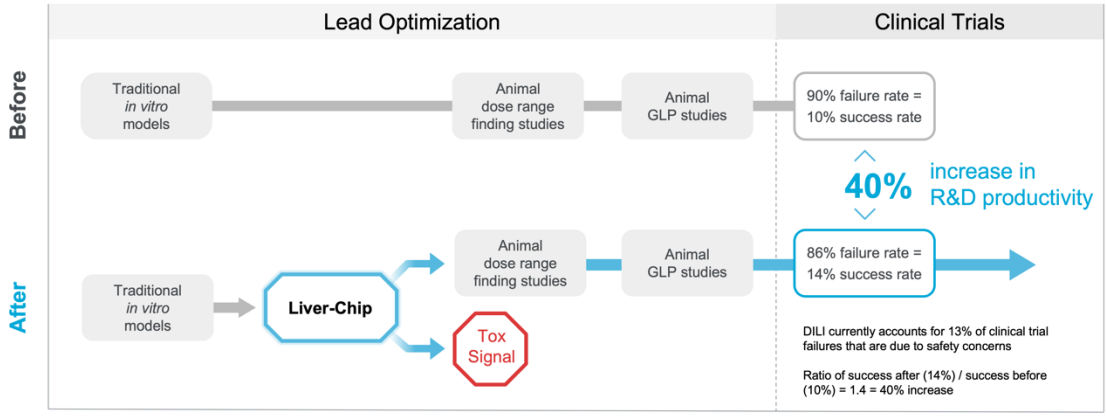

Figure 3. Impact of adding the Liver-Chip to the lead optimization phase of a small molecule drug project Drug projects that do not utilize the Liver-Chip currently proceed from traditional *in vitro* models into *in vivo* dose range finding studies prior to the GLP safety study required for regulatory submission to seek approval to initiate clinical trial. This approach results in only a 10% success rate. With the inclusion of Liver-Chip prior to the *in vivo* dose range finding study, toxic small molecules are more likely to be identified and can be terminated before animal studies begin. In turn this increases the likelihood of a drug candidate being successful in the clinic. The higher success rate translates into a 40% increase in R&D productivity.

Based on this experience, we believe that the Liver-Chip could be employed in the drug development pipeline during the lead optimization phase where projects have identified three-to-five chemical compounds that have the potential to become the candidate drug (Figure 3). If data emerge showing that a chemical compound produces a toxic signal in the Liver-Chip, this will indicate to toxicologists that there is a high (~80%) probability that the compound would similarly cause toxicity in humans. This, in turn, would enable scientists to deprioritize these compounds from early *in vivo* toxicology studies (such as the maximum tolerated dose/dose range finding study) and, consequently, reduce animal usage and advance the "fail early, fail fast" strategy. Importantly, the absence of false positives strengthens the argument that the Liver-Chips should also be adopted within the early discovery phase, as stopping drug candidates that are falsely determined to be toxic by less-robust preclinical models could result in good therapeutics never reaching patients.

Despite these positive findings, it should be acknowledged that the current chip material (PDMS) used in the Liver-Chip may be problematic for a subset of small molecules that are prone to non-specific binding. Although this study demonstrates that the material binding issue does not in practice significantly reduce the predictive value of the Liver-Chip DILI model, work is currently underway to develop chips using materials that have a lower binding potential. It should also be recognized that many pharmaceutical companies have diversified portfolios, with small molecules forming approximately 50% of the therapeutic modalities being developed. Consequently, further investigation of the Liver-Chip performance against large molecules and biologic therapies should be carried out. Integration of resident and circulating immune cells should add even greater predictive capability.

Finally, predictive models that demonstrate concordance with clinical outcomes should provide scientists and corporate leadership with greater confidence in decision-making at major investment milestones. Impressively, our economic analysis revealed $3 billion in improved R&D productivity that could be generated by replacing or supplementing existing preclinical *in vitro* models with human Liver-Chips for this single context of use (DILI prediction). Moreover, an additional productivity value of $24 billion could be gained if

a similar approach is used to develop predictive models for the other most common toxicities that result in drug attrition. Taken together, these results suggest that Organ-Chip technology has great potential to benefit drug development, improve patient safety, and enhance pharmaceutical industry productivity and capital efficiency. This work also provides a starting point for other groups that hope to validate their MPS models for integration into commercial drug pipelines.

Methods

Cell Culture

Cryopreserved primary human hepatocytes, purchased from Gibco (Thermo Fisher Scientific), and cryopreserved primary human liver sinusoidal endothelial cells (LSECs), purchased from Cell Systems, were cultured according to their respective vendor/Emulate protocols. The LSECs were expanded at a 1:1 ratio in 10-15 T-75 flasks (Corning) that were pre-treated with 5mL of Attachment Factor (Cell Systems). Complete LSEC medium includes Cell Systems medium with final concentrations of 1% Pen/Strep (Sigma), 2% CultureBoost (Cell Systems), and 10% Fetal Bovine Serum (FBS) (Sigma). Media was refreshed daily until cells were ready for use. Cryopreserved human Kupffer cells (Samsara Sciences) and human Stellate cells (IXCells) were thawed according to their respective vendor/Emulate protocols on the day of seeding. See Supplementary Table S1 for further information.

Liver Chip Microfabrication and Zoë® Culture Module

Each chip is made from flexible polydimethylsiloxane (PDMS), a transparent viscoelastic polymer material. The chip compartmental chambers consist of two parallel microchannels that are separated by a porous membrane containing pores of 7µm diameter spaced 40µm apart.

On Day -6, chips were functionalized using Emulate proprietary reagents, ER-1 (Emulate reagent: 10461) and ER-2 (Emulate reagent: 10462), mixed at a concentration of 1 mg/mL prior to application to the microfluidic channels of the chip. The platform is then irradiated with high power UV light (peak wavelength: 365nm, intensity: 100 $\mu J/cm^2$) for 20min using a UV oven (CL-1000 Ultraviolet Crosslinker AnalytiK-Jena: 95-0228-01). Chips were then coated with 100 µg/mL Collagen I (Corning) and 25 µg/mL Fibronectin (ThermoFisher) for both channels. The top channel was seeded with primary human hepatocytes on Day -5 at a density of 3.5 x 10^6 cells/mL. Complete hepatocyte seeding medium contains Williams' Medium E (Sigma) with final concentrations of 1% Pen/Strep (Sigma), 1% L-GlutaMAX (Gibco), 1% Insulin-Transferring-Selenium (Gibco), 0.05 mg/mL Ascorbic Acid (Sigma), 1µM dexamethasone (Sigma), and 5% FBS (Sigma). After four hours of attachment, the chips were washed by gravitational force. Gravity wash consisted of gently pipetting 200µL fresh medium at the top inlet, allowing it to flow through, washing out any unbound cells from the surface, and inserting a pipette tip on the outlet of the channel.

On Day -4, a hepatocyte Matrigel overlay procedure was executed with the purpose of promoting a three-dimensional matrix for the hepatocytes to grow in an ECM sandwich culture. The hepatocyte overlay and maintenance medium contain Williams' Medium E (Sigma) with final concentrations of 1% Pen/Strep (Sigma), 1% L-GlutaMAX (Gibco), 1% Insulin-Transferrin-Selenium (Gibco), 50 µg/mL Ascorbic Acid (Sigma), and 100nM Dexamethasone (Sigma). On Day -3, the bottom channel was seeded with LSECs, Stellate cells and Kupffer cells, further known as non-parenchymal cells (NPCs). NPC seeding medium contains Williams' Medium E (Sigma) with final concentrations of 1% Pen/Strep (Sigma), 1% L-GlutaMAX (Gibco), 1% Insulin-Transferrin-Selenium (Gibco), 50 µg/mL Ascorbic Acid (Sigma), and 10% FBS (Sigma). LSECs were detached from flasks using Trypsin (Sigma) and collected accordingly. These cells were seeded in a mixture volume ratio of 1:1:1 with LSECs at a density range of 9-12 x 10^6 cells/mL, Stellates at a density of 0.3 x 10^6 cells/mL, and Kupffer at 6 x 10^6 cells/mL followed by a gravity wash four hours post-seeding.

On Day -2, chips were visually inspected under the ECHO microscope (Discover Echo, Inc.) for cellular maturation and attachment, healthy morphology, and a tight monolayer. The chips that passed visual inspection had both channels washed with their respective media, leaving a droplet on

top. NPC maintenance media was composed of the same components prior, with a reduction of FBS to 2%. To minimize bubbles within the system, one liter of complete, warmed top and bottom media was added to Steriflip-connected tubes (Millipore) in the biosafety cabinet. All media was then degassed using a -70 kPa vacuum (Welch) and stored in the incubator until use. Pods were primed twice with 3mL of degassed media in both inlets, and 200µL in both outlets. Chips were then connected to pods via liquid-to-liquid connection. Chips and pods were placed in the Zoë® for their first regulate cycle, which minimizes bubbles within the fluidic system by increasing the pressure for two hours. After this, normal flow resumed at 30µL/h. On Day -1, Zoës® were set to regulate once more.

Experimental Setup

The 780-chip experiment was carried out in four consecutive cycles (herein referred to as Cycles 1 through 4) to test a selection of 27 drugs at varying concentrations relative to the average therapeutic human C_{max} obtained from literature (Supplementary Table S2). Each cycle tested 6-8 concentrations in duplicate for each of 10-13 drugs. To determine which sampling strategy was optimal for this test plan (16 doses x 1 replicate, 8x2, 5x3, or 4x4), we generated 3 different "dose-response" synthetic datasets, each distorted by different noise levels (low, medium, or high). For each of these datasets, we performed curve-fitting analysis and calculated the Root Mean Square Error between the "true" and estimated IC_{50} parameter. The analysis results showed that, for all noise levels, the 16x1 was the optimal selection and marginally outperformed the 8x2. However, to ensure at least two replicates per concentration, the 8x2 strategy was selected. Some drugs were also repeated across cycles (on different donors or at different concentrations) to ensure experimental robustness. For each cycle, chips were dosed with drug over 8 days (referred to as Day 0 through Day 7). Drug preparation, dosing, and analysis teams were divided, creating a double-blind study such that those administering the drugs or performing analyses did not know the name or concentrations of the drugs tested.

Drug Preparation

The drug dosing concentrations were determined from the unbound human C_{max} of each drug. First, the expected fraction of drug unbound in media with 2% FBS was extrapolated from plasma binding data for each drug. Dosing concentrations were then back calculated such that the unbound fraction in media would reflect relevant multiples of unbound human C_{max} (Supplementary Table S2). For each cycle, concentrations ranged from 0.1 to 1000 times C_{max}.

Stock solutions were prepared at 1000 times the final dosing concentration. Drugs in powder form were either weighed out with 1mg precision or dissolved directly in vendor-provided vials. Sterile DMSO (Sigma) was added to dissolve the drug. The solution was triturated before transferring to an amber vial (Qorpak), which was vortexed (Fisher Scientific) on high for 60 seconds to ensure complete dissolution. A serial dilution was then performed in DMSO to prepare each subsequent 1000X concentration. These stock solutions were then aliquoted in 1.5mL tubes (Eppendorf) and stored at -20˚C until dosing day, allowing a maximum of one freeze-thaw cycle prior to dosing.

All media was made the day prior to chip dosing and stored overnight at 37˚C. On the day of chip dosing, one stock aliquot per drug concentration was thawed in a 37˚C bead bath. The stocks were then vortexed and inspected to ensure absence of drug particulate. Dosing solutions were prepared by diluting drug stock 1:1000 in top or bottom media to achieve 0.1% DMSO concentration. The dosing solutions were then vortexed and stored at 37˚C until dosing.

On the first dosing day (Day 0), all chips were imaged using the ECHO microscope. 500µL of effluent was collected from all four reservoirs of the pod and placed in a labelled 96-well plate. After collection, all the remaining media was carefully aspirated before dosing with 3.8mL of corresponding dosing solution. Dosing occurred on study days 0, 2, and 4 for chips flowing at 30 µL/h, and on days 0, 1, 2, 3, 4, 5, and 6 for chips flowing at 150 µL/h. Effluent collection occurred on days 1, 3, and 7.

Compound Distribution Kit

Prior to the study, each drug was classified based on its partition coefficient (logP), polar surface area, and molecular weight to determine the risk of absorption into the PDMS during channel perfusion. Drug absorption and distribution, which determines the concentration at which cells are truly exposed, was factored into the study design by assigning drugs classified as higher risk to Cycle 2 of the study. Prior to Cycle 2, a compound distribution kit (CDK) experiment (EP129 v1.0) was completed to examine the levels at which these specific drugs would be absorbed into the PDMS system. Dosing solutions at the highest concentration for each drug of interest were prepared and perfused through both channels (absent ECM or cells), with inlet and outlet reservoir samples taken at days 1 and 7.

The samples were shipped to Charles River Laboratories (CRL, Worcester, MA) for liquid-chromatography-mass-spectrometry (LC/MS) analysis, wherein the ratios of inlet drug concentration to outlet effluent concentration were calculated. Buspirone, chlorpheniramine, nefazodone, and simvastatin were identified as being highly absorbed into PDMS. These four drugs were perfused at 150µL/h in Cycle 2, pursuant to the Emulate-validated solution to reduce absorption.

LC/MS Sample Collection

LC/MS analysis was also performed on inlet and outlet effluent samples from each chip at 0.3x, 3x, 30x, and 300x concentrations. 100µL of effluent was collected and plated in a 96-well plate, which was stored at -80˚C until shipped directly over dry ice via courier. In Cycle 2, the plates underwent one additional freeze/thaw cycle (thawed for 5 hours at 4˚C) for logistical purposes. On dosing days, dosing solution was also collected for LC/MS analysis to determine true dosing concentration and quantify variability in the dosing protocol.

LC/MS Analysis

Upon arrival at Charles River Laboratories, plates were stored at -80˚C until analysis. Samples were processed via protein precipitation (PPT) extraction method. Briefly, stock solutions were prepared at 10mM in DMSO (with the exception of diclofenac and asunaprevir prepared at 1mM). 10µL stock solution (blank matrix for control) was combined with 60µL internal standard (acetonitrile for control blanks) before vortexing and centrifuging. All samples were combined with 50µL MilliQ Water before analysis. Liquid chromatography analysis used Waters HSS T3 column (2.5µM; 50x2.1mm) at 55˚C. Mass spectrometry analysis used electrospray ionization method to develop positive ions at 550˚C, with <0.1% carryover.

Biochemical Assays

Top channel outlet effluents were analyzed to quantify albumin and alanine transaminase (ALT) levels on days 1, 3, and 7, using sandwich ELISA kits (Abcam, Albumin ab179887, ALT ab234578), according to vendor-provided protocols. Frozen (-80˚C) effluent samples were thawed overnight at 4˚C prior to assay. The Hamilton Vantage liquid handling platform was used to manage effluent dilutions (1:500 for albumin, neat for ALT), preparation of standard curves, and addition of antibody cocktail. Absorbance at 450nm was measured using the SynergyNeo Microplate Reader (BioTek).

As part of Cycle 3, top channel outlet samples were analyzed to quantify urea levels with a urea assay kit (Sigma-Aldrich, MAK006) according to vendor-provided protocol. Frozen (-80˚C) effluent samples were thawed overnight at 4˚C prior to assay. Samples of vehicle chips from days 1, 3, and 7 were analyzed, as well as levofloxacin, trovafloxacin, pioglitazone, and troglitazone on day 3. All samples were diluted 1:5 in assay buffer and mixed with the kit's Reaction Mix. Absorbance at 570nm was measured using the same automated plate reader.

Morphology Analysis

At least four to six brightfield images were acquired per chip for morphology analysis. Brightfield images were acquired on the ECHO microscope using these settings: 170% zoomed phase contrast, 50% LED, 38% brightness, 41% contrast, 50% color balance, color on, and 10X objective. Brightfield images were acquired across three fields of view on days 1, 3, and 7 for each cycle.

Cytotoxicity classification was performed while acquiring images for both NPCs and hepatocytes. Images were then scored zero to four by blinded individuals (n=2) based on severity of agglomeration of cell debris for both channels.

At the end of the experiment, cells in the Liver-Chip were fixed using 4% paraformaldehyde (PFA) solution (Electron Microscopy Sciences). Chips were detached from pods and washed once with PBS. The PFA solution was pipetted into both channels and incubated for 20 minutes at room temperature. Afterwards, chips were washed with PBS and stored at 4˚C until staining.

Fixed Staining

Following fixation, chips corresponding to low, medium, and high concentrations from each group were cut in half with a razor blade perpendicular to the co-culture channels. One half was used in the following staining protocol, while the other half was stored for future staining. All stains and washes utilized the bubble method, in which a small amount of air is flowed through the channel prior to bulk wash media to prevent a liquid-liquid dilution of the staining solution. The top channel was perfused with 100µL of NucBlue (ThermoFisher, R37605) (100 drops in 50mL of PBS) to visualize cell nuclei. Following 15 minutes of incubation at room temperature, each channel was washed with 200µL of PBS (alternating channels, 2x for top and 3x for bottom). Chips were then imaged using the Opera Phenix Plus.

Following DAPI staining and imaging, chips were stained with a multi-compound resistant protein 2 (MRP2) antibody in order to visualize the bile canalicular structures characteristic of healthy Liver-Chips. First, chips were permeabilized in 0.125% Triton-X and 2% Normal Donkey Serum (NDS) diluted in PBS (100µL of solution per channel) and incubated at room temperature for 10 minutes. Then, each channel was washed with 200µL of PBS (alternating channels, 2x for top and 3x for bottom). Chips were then blocked in 2% Bovine Serum Albumin (BSA) and 10% NDS in PBS (100µL of solution per channel) and incubated at room temperature for 1 hour. Next, primary antibody Mouse anti-MRP2 (Abcam, ab3373) was prepared 1:100 in the original blocking buffer, diluted 1:4 in PBS. 100µL of solution was added to each channel, and chips were stored overnight at 4˚C. The following day, each channel was washed with 200µL of PBS (alternating channels, 2x for top and 3x for bottom). Secondary antibody Donkey anti-Mouse 647 (Abcam, ab150107) was prepared 1:500 in original blocking buffer, diluted 1:4 in PBS. 100µL of solution was added to each channel and chips incubated at room temperature, protected from light, for two hours. Then each channel was washed with 200µL of PBS (alternating channels, 2x for top and 3x for bottom). Chips were imaged immediately or stored at 4˚C until ready for imaging on the Opera Phenix Plus.

Live Staining

Chip replicates designated for live cell imaging were washed with PBS utilizing the bubble method. Chips were then cut in half perpendicular to the co-culture channels. The top chip halves were stained with NucBlue (ThermoFisher, R37-605) to visualize cell nuclei and Cell Event Green (ThermoFisher, C10423) to visualize activated caspase-3/7 for apoptosis. This staining panel was prepared in serum-free media (CSC), with NucBlue at 2 drops per mL and Cell Event Green at a 1:500 ratio and perfused through both channels. The bottom chips halves were stained with NucBlue (Thermo) to visualize nuclei and Tetramethylrhodamine, methyl ester (TMRM) (ThermoFisher, I34361) to visualize active mitochondria. This staining panel was prepared in PBS with 5% FBS, with NucBlue at 2 drops per mL and TMRM at a 1:1000 ratio in original blocking buffer, diluted 1:4 in PBS. Chips were incubated in the dark at 37˚C for 30 minutes, and then each channel was washed with 200µL of PBS (alternating channels, 2x for top and 3x for bottom). The chips were kept at 37˚C, protected from light, until ready for imaging with the Opera Phenix Plus.

Image Acquisition

Fluorescent confocal image acquisition was performed using the Opera Phenix Plus High-Content Screening System and Harmony 4.9 Imaging and Analysis Software (PerkinElmer). Before acquisition, the Phenix internal environment was set to 37˚C and 5% CO_2. Chips designated for imaging were removed from their plates, wiped on the

bottom surface to remove moisture, and placed into the Phenix 12-chip imaging adapter. Whole chips were placed directly into each slot, while top and bottom half chips were matched and combined in one chip slot. Chips were aligned flush with the adapter and one another. Any bubbles identified from visual inspection were washed out with PBS. Once ready, the stained chips were covered with transparent plate film to seal channel ports and loaded into the Phenix. For live imaging, the DAPI (Time: 200ms, Power: 100%), Alexa 488 (Time: 100ms, Power: 100%), and TRITC (Time: 100ms, Power: 100%) lasers were used. For fixed imaging, the DAPI (Time: 200ms, Power: 100%), TRITC (Time: 100ms, Power: 100%), and Alexa 647 (Time: 300ms, Power: 80%) lasers were used. Z-stacks were generated with 3.6µm between slices for 28-32 planes, so that the epithelium was located around the center of the stack. Six fields of view (FOVs) per chip were acquired, with a 5% overlap between adjacent FOVs to generate a global overlay view.

Image Analysis

Raw images from fixed and live imaging were exported In TIFF format from the Harmony software. Using scripts written for FIJI (ImageJ), TIFFs across 3 color channels and multiple z-stacks were compiled into composite images for each field of view in each chip. The epithelial signal was identified and isolated from the endothelial and membrane signals, and the composite TIFFs were split accordingly. The ideal threshold intensity for each channel in the epithelial "substack" was identified to maximize signal, and the TIFFs were exported as JPEGs for further analysis.

Gene Expression Analysis

RNA was extracted from chips using TRI Reagent (Sigma-Aldrich) according to the manufacturer's guidelines. The collected samples were submitted to GENEWIZ (South Plainfield, NJ) for next-generation sequencing. After quality control and RNA-seq library preparation, the samples were sequenced with the Illumina HiSeq 4000 at 2x150bp using sequencing depth of ~50M paired-end reads/sample. Using Trimmomatic v0.36, the sequence reads were trimmed to filter out all poor-quality nucleotides and possible adapter sequences. The remaining trimmed reads were mapped to the Homo sapiens reference genome GRCh38 using the STAR aligner v2.5.2b. Next, using the generated BAM files for each sample, the unique gene hit counts were calculated from the Subread package v1.5.2. It is worth noting that only unique reads within the exon region were counted. Finally, the generated hit-counts were used to perform DGE analysis using the "DESeq2" R package[54]. The thresholds for all the DGE analyses were: $|\log_2(\text{Fold Change})| \geq 1$ and adjusted p-value ≤ 1.

Statistical Analysis

All statistical analyses were conducted in R[70] (version 4.1.2), and figures were produced using the R package ggplot2[71] (version 3.3.5). The dose-response analysis (Figure 2a) was carried out using the popular drc R package developed by Ritz et al.[72] using the generalized log-logistic dose response model. The error bars in Figures 1d, 1e and 2a correspond to the standard errors of the mean. The circles in Figure 1e correspond to the samples used to calculate the corresponding statistics. In Figures 1e and 2a the number of samples used were n=4 and n=2 respectively. Finally, the analysis of significance in Figures 1d and 1e was performed using unpaired and paired t-test respectively (P-value < 0.05).

Acknowledgements

The authors would like to thank Dr. Jack Scannell for his invaluable guidance on study design and data interpretation. We would also like to recognize the wider Emulate team for their valuable discussion and support. Finally, we are grateful for Aaron VanDevender's critical evaluation and review of the manuscript.

Author contributions

L.E., D.E.I., D.L., D.V.M., J.D.S. designed the study; A.A., S.A.B., J.T.C., C.V.C., A.R.H., J.J., S.J., K.K.M., M.E.Q., A.C.R., W.T., M.W., J.V., contributed to drug preparation, execution of the study, downstream processing of images and data collection; J.J., S.J., G.K., V.J.K., C.Y.L., C. L., J.S.R., D.R.T., M.W., chip seeding, maintenance, drug administration and chip sampling; C.V.C., S.J., G.K., V.J.K., provided leadership for one full

experimental cycle; L.E., C.L., D.V.M., S.J., M.K., J.D.S. contributed to the data analysis; A.R.H., L.E., D.E.I., C.Y.L., K.K.M., M.W. contributed to writing the manuscript; K-J. J., O.I., P.K.M. performed a blinded data review and critically reviewed the manuscript.

Competing interests

L.E., D.L., D.V.M., J.D.S., A.A., S.A.B., J.T.C., C.V.C., A.R.H., J.J., S.J., M.K., K.K.M., M.E.Q., A.C.R., W.T., M.W., G.K., V.J.K., C.Y.L., C. L., J.S.R., D.R.T., J.V., K-J.J. are employees of Emulate and may hold equity; D.E.I. is a founder, board member, SAB chair, and equity holder in Emulate Inc.

References

1. Khanna, I. Drug discovery in pharmaceutical industry: Productivity challenges and trends. Drug Discov. Today 17, 1088–1102 (2012).
2. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? Nat. Rev. Drug Discov. 3, 711–715 (2004).
3. Peck, R. W., Lendrem, D. W., Grant, I., Lendrem, B. C. & Isaacs, J. D. Why is it hard to terminate failing projects in pharmaceutical R&D? Nat. Rev. Drug Discov. 14, 663–664 (2015).
4. Scannell, J. W. & Bosley, J. When Quality Beats Quantity: Decision Theory, Drug Discovery, and the Reproducibility Crisis. PLoS One 11, e0147215 (2016).
5. Paul, S. M. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat. Rev. Drug Discov. 2010 93 9, 203–214 (2010).
6. Cook, D. et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat. Rev. Drug Discov. 13, 419–431 (2014).
7. Morgan, P. et al. Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nat. Rev. Drug Discov. 2018 173 17, 167–181 (2018).
8. Weaver, R. J. et al. Managing the Challenge of Drug Induced Liver Injury: A roadmap for the development and deployment of preclinical predictive models. 1–61 (2019).
9. Wu, F. et al. Computational Approaches in Preclinical Studies on Drug Discovery and Development. Front. Chem. 8, 726 (2020).
10. Ferreira, G. S. et al. A standardised framework to identify optimal animal models for efficacy assessment in drug development. PLoS One 14, (2019).
11. Bhatia, S. N. & Ingber, D. E. Microfluidic organs-on-chips. Nat. Biotechnol. 2014 328 32, 760–772 (2014).
12. Esch, E. W. et al. Organs-on-chips at the frontiers of drug discovery. Nat. Rev. Drug Discov. 14, 248–260 (2015).
13. Huh, D. et al. Reconstituting organ-level lung functions on a chip. Science 328, 1662–1668 (2010).
14. Kasendra, M. et al. Duodenum intestine-chip for preclinical drug assessment in a relevant human model. Elife 9, (2020).
15. Kerns, S. J. et al. Human immunocompetent Organ-on-Chip platforms allow safety profiling of tumor-targeted T-cell bispecific antibodies. Elife 10, (2021).
16. Jalili-Firoozinezhad, S. et al. A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. Nat. Biomed. Eng. 2019 37 3, 520–531 (2019).
17. Chou, D. B. et al. On-chip recapitulation of clinical bone marrow toxicities and patient-specific pathophysiology. Nat. Biomed. Eng. 2020 44 4, 394–406 (2020).
18. Fabre, K. et al. Introduction to a manuscript series on the characterization and use of microphysiological systems (MPS) in pharmaceutical safety and ADME applications. Lab Chip 20, 1049–1057 (2020).
19. Baudy, A. R. et al. Liver microphysiological systems development guidelines for safety risk assessment in the pharmaceutical industry. Lab on a Chip vol. 20 215–225 (2020).
20. Zhou, Y., Shen, J. X. & Lauschke, V. M. Comprehensive evaluation of organotypic and microphysiological liver models for prediction of drug-induced liver injury. Front. Pharmacol. 10, 1–22 (2019).

21. Jang, K.-J. et al. Reproducing human and cross-species drug toxicities using a Liver-Chip. Sci. Transl. Med vol. 11 http://stm.sciencemag.org/ (2019).
22. Ribeiro, A. J. S., Yang, X., Patel, V., Madabushi, R. & Strauss, D. G. Liver Microphysiological Systems for Predicting and Evaluating Drug Effects. Clin. Pharmacol. Ther. 106, 139–147 (2019).
23. Rodríguez-Antona, C. et al. Cytochrome P450 expression in human hepatocytes and hepatoma cell lines: molecular mechanisms that determine lower expression in cultured cells. Xenobiotica. 32, 505–520 (2002).
24. Inada, H. et al. Direct reprogramming of human umbilical vein- and peripheral blood-derived endothelial cells into hepatic progenitor cells. Nat. Commun. 2020 111 11, 1–17 (2020).
25. Foster, A. J. et al. Integrated in vitro models for hepatic safety and metabolism: evaluation of a human Liver-Chip and liver spheroid. Arch. Toxicol. 93, (2019).
26. Proctor, W. R. et al. Utility of spherical human liver microtissues for prediction of clinical drug-induced liver injury. Arch. Toxicol. 91, 2849–2863 (2017).
27. Vorrink, S. U., Zhou, Y., Ingelman-Sundberg, M. & Lauschke, V. M. Prediction of drug-induced hepatotoxicity using long-term stable primary hepatic 3D spheroid cultures in chemically defined conditions. Toxicol. Sci. 163, 655–665 (2018).
28. O'Brien, P. J. et al. High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. 580–604 (2006) doi:10.1007/s00204-006-0091-3.
29. Godoy, P. et al. Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME. Archives of Toxicology vol. 87 (2013).
30. Heise, T. et al. Insulin Degludec: Four Times Lower Pharmacodynamic Variability Than Insulin Glargine Under Steady-State Conditions in Type 1 Diabetes. 1–6 (2012).
31. Waldmann, T. et al. Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. Chem. Res. Toxicol. 27, 408–420 (2014).
32. Albrecht, W., Kappenberg, F., Brecklinghaus, T., Stoeber, R. & Marchan, R. Prediction of human drug induced liver injury (DILI) in relation to oral doses and blood concentrations. Arch. Toxicol. 93, 1609–1637 (2019).
33. Xu, J. J. et al. Cellular imaging predictions of clinical drug-induced liver injury. Tox. Sci. 105, 97–105 (2008).
34. Shaw, P. J., Ganey, P. E. & Roth, R. A. Idiosyncratic Drug-Induced Liver Injury and the Role of Inflammatory Stress with an Emphasis on an Animal Model of Trovafloxacin Hepatotoxicity. Toxicol. Sci. 118, 7–18 (2010).
35. Castiella, A., Zapata, E., Lucena, I., Andrade, R. J. & Service, G. drug-induced autoimmune liver disease: a diagnostic dilemma of an increasingly reported disease. Available from World J Hepatol 6, 160–168 (2014).
36. Lomitapide. LiverTox Clin. Res. Inf. Drug-Induced Liver Inj. (2019).
37. Joshi, E. M., Heasley, B. H., Chordia, M. D. & Macdonald, T. L. In vitro metabolism of 2-acetylbenzothiophene: relevance to zileuton hepatotoxicity. Chem. Res. Toxicol. 17, 137–143 (2004).
38. Machinist, J. M., Kukulka, M. J. & Bopp, B. A. In Vitro Plasma Protein Binding of Zileuton and its N-Dehydroxylated Metabolite. Clin. Pharmacokinet. 1995 292 29, 34–41 (2012).
39. Hendriks, D. F. G. et al. Mechanisms of Chronic Fialuridine Hepatotoxicity as Revealed in Primary Human Hepatocyte Spheroids. Toxicol. Sci. 171, 385-395 (2019).
40. Li, F., Cao, L., Parikh, S. & Zuo, R. Three-Dimensional Spheroids With Primary Human Liver Cells and Differential Roles of Kupffer Cells in Drug-Induced Liver Injury. J. Pharm. Sci. 109, 1912-1923 (2020).

41. Simon, T. et al. Combined glutathione-S-transferase M1 and T1 genetic polymorphism and tacrine hepatotoxicity. Clin. Pharmacol. Ther. 67, 432–437 (2000).
42. Garside, H. et al. Evaluation of the use of imaging parameters for the detection of compound-induced hepatotoxicity in 384-well cultures of HepG2 cells and cryopreserved primary human hepatocytes. Tox. In Vitro 28, 171–181 (2014).
43. Levoquin (Levofloxacin) Product Monogram, Jansen Inc., Toronto, Ontario, M3C IL9, Revised July 2011.
44. J. B. Kahn, Latest Industry Information on the Safety Profile of Levofloxacin in the US, Chemotherapy, 47, Suppl. 3, 32-37 (2001).
45. Godoy, P. et al. Extracellular matrix modulates sensitivity of hepatocytes to fibroblastoid dedifferentiation and transforming growth factor β-induced apoptosis. Hepatology 49, 2031–2043 (2009).
46. Redfern, W. S. et al. Impact and frequency of different toxicities throughout the pharmaceutical life cycle. Toxicol. 114, 231 (2010).
47. Low, L. A. et al. Organs-on-chips: into the next decade. Nat. Rev. Drug Discov. 20, 345–361 (2021).
48. Roth, A. et al. Human microphysiological systems for drug development. Science, 373, 1304-1306 (2021).
49. Scannell, J. W. et al. Diagnosing the decline in pharmaceutical R&D efficiency. Nat. Rev. Drug Discov. 11, 191–200 (2012).
50. Hornberg, J. J. et al. Exploratory toxicology as an integrated part of drug discovery. Part II: Screening strategies. Drug Discov. Today 19, 1137–1144 (2014).
51. Eckert, J. E. et al. Recommended Guidelines for Developing, Qualifying, and Implementing Complex In Vitro Models (CIVMs) for Drug Discovery. SLAS Discov. 25, 1174–1190 (2020).
52. Mahony, C. et al. New ideas for non-animal approaches to predict repeated-dose systemic toxicity: Report from an EPAA Blue Sky Workshop. Reg. Tox. Pharm. 114, 104668 (2020).
53. Dourson, M. et al. The Future of Uncertainity Factors with in vitro Studies using human cells. Tox Sci. 1-6 (2021).
54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 1-21 (2014).
55. Bristol-Myers Squibb. SUNVEPRA® [prescribing information]. 1, 1–22 (2017).
56. Smith, G., Goulbourn, R., Burt, R. & Chatfield, D. Preliminary studies of absorption and excretion of benoxaprofen in man. Br. J. Clin. Pharmacol. 4, 585–590 (1977).
57. Porceddu, M. et al. Prediction of liver injury induced by chemicals in human with a multiparametric assay on isolated mouse liver mitochondria. Toxicol. Sci. 129, 332–345 (2012).
58. Regenthal, R. & Koeppel, C. Drug Levels: Therapeutic and Toxic Serum/Plasma Concentrations of Common Drugs. J. Clin. Monit. Comput. 15, 529–544 (1999).
59. Schulz, M. & Schmoldt, A. Therapeutic and toxic blood concentrations of more than 800 drugs and other xenobiotics. Pharmazie 58, 447–474 (2003).
60. Lau, S.W.J. Center for Drug Evaluation and Research Application Number: 203858Orig1s000 Clinical Pharmacology and Biopharmaceutics Review(s). (2012).
61. Chatsiricharoenkul, S., Niyomnaitham, S., Pongnarin, P., Sathirakul, K. & Kongpatanakul, S. Bioequivalence study of 10 mg olanzapine tablets in healthy Thai volunteers. J. Bioequivalence Bioavailab. 3, 082–085 (2011).
62. Scott, L. J. ADIS DRUG PROFILE Sitaxentan In Pulmonary Arterial Hypertension. Drugs vol. 67 (2007).
63. Sivapalasingam, S. & Steigbigel, N. H. Macrolides, Clindamycin, and Ketolides. Mand. Douglas, Bennett's Princ. Pract. Infect. Dis. 1, 358–376 (2015).
64. National Center for Biotechnology Information. PubChem Compound Summary for CID 6918493, Ambrisentan. (2021).
65. Lombardo, F., Berellini, G. & Obach, R. S. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 1352 drug compounds. Drug Metab. Dispos. 46, 1466–1477 (2018).

66. Dawson, S., Stahl, S., Paul, N., Barber, J. & Kenna, J. G. In vitro inhibition of the bile salt export pump correlates with risk of cholestatic drug-induced liver injury in humans. Drug Metab. Dispos. 40, 130–138 (2012).
67. Brunton, L. L., Hilal-Dandan, R. & Knollmann, B. C. Goodman & Gilman's: The Pharmacological Basis of Therapeutics. (2018).
68. National Center for Biotechnology Information. PubChem Compound Summary for CID 216235, Sitaxentan. (2021).
69. Shibukawa, A., Sawada, T., Nakao, C., Izumi, T. & Nakagawa, T. High-performance frontal analysis for the study of protein binding of troglitazone (CS-045) in albumin solution and in human plasma. J. Chromatogr. A 697, 337–343 (1995).
70. R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2021).
71. Wickham, H. Ggplot2: Elegant Graphics for Data Analysis. 2nd Edition, Springer, New York. https://doi.org/10.1007/978-0-387-98141-3 (2009).
72. Ritz C., Baty F., Streibig J.C., Gerhard D. Dose-Response Analysis Using R. PLOS ONE 10(12): e0146021. https://doi.org/10.1371/journal.pone.0146021 (2015)

References 54 to 72 relate to supplementary material