Innovative Organotypic *In Vitro* Models for Safety Assessment:

Aligning with regulatory requirements and understanding models of the heart, skin and liver as paradigms

**Chris S. Pridgeon 1, Constanze Schlott 1, Min Wei Wong 1, Minne B. Heringa 2, Tobias Heckel 3, Joe Leedale 4 Laurence Launay 5, Vitalina Gryshkova 6, Stefan Przyborski 7, Rachel N. Bearon 4, Emma L. Wilkinson 1, Tahera Ansari 8, John Greenman 9, Delilah F. G. Hendriks 10, Sue Gibbs 11, James Sidaway 12, Rowena L. Sison-Young 1, Paul Walker 13, Mike J. Cross 1, B. Kevin Park 1, Chris E. P. Goldring 1**

**1 Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, L69 3GE, UK**

**2 National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands; written in personal capacity.**

**3 Dr. Johannes Heidenhain GmbH, Dr.-Johannes-Heidenhain-Straße 5, D-83301 Traunreut, Germany**

**4** Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK

**5 Servier, Paris, France**

**6 Investigative Toxicology, Department of Non-Clinical Development, UCB Biopharma SPRL, B-1420 Braine L’Alleud, Belgium, Vitalina.Gryshkova@ucb.com**

**7** Department of Biosciences, Durham University, Durham, DH1 3LE, UK

**8 Northwick Park Institute for Medical Research, Northwick Park and St Mark’s Hospital, Middlesex, HA1 3UJ, UK**

**9 School of Life Sciences, University of Hull, Hull, HU6 7RX, UK**

**10 Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden**

**11 Department of Dermatology, VU university medical center; and Department of Oral Cell Biology, Academic Center for Dentistry Amsterdam, University of Amsterdam and VU University, The Netherlands**

**12 Phenotox Ltd, Cheshire, SK10 5QT, UK**

**13 Cyprotex Discovery Ltd, Cheshire, SK10 4TG, UK**

# Abstract

The development of improved, innovative models for the detection of toxicity of drugs, chemicals or chemicals in cosmetics is crucial to efficiently bring new products safely to market in a cost-effective and timely manner. Additionally, improvement in models to detect toxicity may reduce the incidence of unexpected post-marketing toxicity and reduce or eliminate the need for animal testing.

The safety of novel products of the pharmaceutical, chemical or cosmetics industry must be assured, therefore toxicological properties need to be assessed. Accepted methods for gathering the information required by law for approval of substances are often animal methods. To reduce, refine and replace animal testing, innovative organotypic *in vitro* models have emerged. Such models appear at different levels of complexity ranging from simpler, self-organized three-dimensional (3D) cell cultures up to more advanced scaffold-based co-cultures consisting of multiple cell types. This review provides an overview of recent developments in the field of toxicity testing with *in vitro* models for three major organ types: heart, skin, and liver. This review also examines regulatory aspects of such models in Europe and the UK, and summarizes best practices to facilitate the acceptance and appropriate use of advanced *in vitro* models

# Introduction

When developing novel drugs, chemicals or personal care products, industry must evaluate the risks to human health arising from their use. Therefore, knowledge of the properties of these substances, results of safety tests, risk assessments, and appropriate measures to adequately control the risks must be provided to regulatory authorities (Eichler et al., 2008; Pignatti et al., 2011; Senderowicz, 2010; Silbergeld et al., 2015). This information is mandatory for registration and marketing approval as well as for approval of clinical trials for drugs or personal care products.

Traditionally such risk assessments are based on safety tests performed in animals and assume that animals will respond to these tests in a similar manner to humans. Although animals represent systemic organisms with obvious similarities in physiology and function to humans, there are also several limitations. Animal testing is labour-intensive, time-consuming, expensive, ethically challenging and not suited to address the high number of substances produced by the chemical industry or during drug screening in the pharmaceutical industry (Fig 1) (Hartung, 2010; Kessel and Frank, 2007). Furthermore, known species variation makes reliance on tests in a single species insufficient for approval of clinical trials in humans (Zbinden, 1993). This is why regulatory authorities for pharmaceuticals require *in vivo* testing in two species, commonly a rodent such as rat or mouse and a non-rodent such as mini-pig, dog, or cynomolgus monkey (Bode et al., 2010; Greaves et al., 2004).

The burden placed upon animal testing has been a contentious subject for many decades, with organised opposition since the 19th century (Finn and Stark, 2015). Similar ideals are proposed in the EU, UK, US and in particular by the British National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) who follow the principles of humane experimental technique conceptualized by Russell and Burch (Russell et al., 1959). Validated *in vitro* tests for phototoxicity, cytotoxicity, genotoxicity, and cardiac arrhythmias are already well established to test for cell and DNA damage as well as for cardiac ion channel inhibition. These tests are based on simple cell culture systems employing mammalian cells for photo- and cytotoxicity screens, bacteria and mammalian cells for the Ames mutagenicity test and *in vitro* micronucleus test (MNT), or cell lines expressing the human potassium channel hERG in cardiac safety assays (Bridgland-Taylor et al., 2006; Colatsky et al., 2016; Kirkland et al., 2014; Moeller et al., 2012; Shukla et al., 2010; Spielmann et al., 2008).

Testing with *in vitro* assays and animal tests is relatively effective in the detection of acute and severe toxicities and to some degree in testing for chronic toxicities. However, these tests still exhibit limitations, since ~1/3 of candidate drugs fail during clinical trials due to unpredicted toxicity (Arrowsmith and Miller, 2013; Cook et al., 2014). Many of these safety failures can be attributed to cardiovascular and liver toxicities (Fig 1C) (Cook et al., 2014; Laverty et al., 2011). An exemplar case of unexpected toxicity is the clinical trial of the antiviral agent Fialuridine in 1992, in which unforeseen toxicity led to the death of a third of the patient cohort due to liver failure associated with lactic acidosis. Two of the remaining patients required a liver transplant. This toxicity is very pertinent as it was not uncommon in humans, and demonstrates the predictive limitations of pre-clinical studies (McKenzie et al., 1995). More recent examples include the TGN1412 trial, where despite the use of animal studies for the novel immunomodulatory antibody CD28 ‘superagonist’ and use of a dose 500 times lower than found to be safe in animals, all six of the patients suffered from cytokine storm and were hospitalised (Suntharalingam et al., 2006). Additionally, animal studies were not predictive of the toxicity observed in first-in-man trials for the fatty acid amide hydrolase inhibitor BIA 10-2474, where five patients suffered neurological injuries and a sixth died (Moore, 2016). These failures may be due to a lack in predictivity due to the phylogenetic distance between laboratory animals and humans, as well as the discrepancy between simplistic *in vitro* tests and the *in vivo* situation. Animals are not fully predictive of human toxicity and *in vitro* tests on traditional two-dimensional (2D) monolayers of cells are neither physiological nor systemic.

Three-dimensional (3D) cell culture and organotypic *in vitro* models are another approach to bridge the gap between traditional 2D cell culture models and the *in vivo* situation. Three-dimensional culture produces cells with more physiologically relevant attributes, such as cell polarization, cell-cell or cell-microenvironment interactions, lumen formation, reduced proliferation, increased differentiation, and numerous changes in RNA and protein expression (Edmondson et al., 2014; Kenny et al., 2007; Rimann and Graf-Hausner, 2012; Yamada and Cukierman, 2007). Therefore, these models hold promise to better represent the histological and physiological complexity of real tissue to study toxicological effects during product development and life cycle management (Fig 1). The need for innovative models has become particularly urgent for the cosmetics industry following a complete ban on cosmetics developed through animal testing in the European Union since 2013 (EU Regulation No 1223/2009). Consequently, human *in vitro* skin equivalents are probably the most developed and understood *in vitro* engineered 3D model for compound testing (Mathes et al., 2014).

In this review, we discuss innovative *in vitro* models currently being used or recently developed as well as the regulatory perspective for toxicological safety assessments in the pharmaceutical, chemical, or cosmetics industry, in order to draw up recommendations for the way forward.

# Regulation

The limitations of animal testing and 2D *in vitro* systems demonstrate a clear need for better models that can be accepted by regulatory agencies. For example, the EU ban on animal testing for cosmetics the and the ambition of the Dutch government to phase out the use of laboratory animals for regulatory safety testing by 2025 (Netherlands National Committee for the protection of animals used for scientific purposes, 2016) necessitate regulatory acceptance of alternative methods. Alternative cell culture methods have already emerged in the industry for compound screening prior to regulatory testing. A first aspect to consider in regulatory acceptance is whether legal frameworks allow alternative methods.

Toxicological data requirements for the evaluation and admission of chemical substances on the European market are given in 11 European regulatory frameworks. Analyses of these frameworks has revealed that although most frameworks name certain animal tests as standard for providing certain toxicological information (e.g. for repeated dose toxicity), all but one clearly provide for using alternative methods to obtain this information (Heringa et al., 2014; Vonk et al., 2015). The exception is the framework for veterinary medicinal products, where the legal status of such a possibility is unclear, as this is only provided in a non-binding guideline. In summary, there are no legal barriers to omit animal safety tests during the safety assessment of a novel chemical entity.

The possibility to acquire regulatory acceptance for clinical trial applications without animal safety testing was illustrated in 2010 by the biotechnology company Immunocore Ltd, who received approval for clinical trials in melanoma patients with an immunostimulating biological without animal testing (Megit, 2011). In their dialogue with the British Medicines and Healthcare products Regulatory Agency (MHRA) and US Food and Drug Administration (FDA), Immunocore brought the arguments forward that their biological can only bind and show activity with human cells and that a relevant animal model is not available for safety evaluation; and that in this case, animal tests have no value. Therefore it was concluded that extensive testing with human cells and human tissues were sufficient and that toxicity studies in non-relevant species may be misleading and are discouraged, which is in agreement with the International Regulatory Guideline ICH Topic S6: Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals (CPMP/ICH/302/95). The MHRA stresses that deviations from the standard safety package are possible when well-justified, which is evaluated on a ‘case by case’ basis. The agency recommends that companies consult for advice on the appropriateness of their development programmes before conducting unnecessary and potentially misleading studies. The MHRA therefore offers a ‘safe harbour’ approach to researchers presenting new models who may otherwise be concerned about punitive action from regulators. Indeed, where new models are being developed, regulators should be viewed as an ally rather than a hindrance to progress.

This, and the legal situation, is encouraging for the application of alternative models, but raises the question as to why animal testing for safety evaluations have not yet been replaced. The main reason is the lack of scientifically acceptable and physiologically-relevant alternatives for animal safety tests. Before approval, a thorough validation is necessary, ideally followed by acceptance within the OECD to achieve mutual recognition of data worldwide.

The validation of alternative tests is the process by which the reliability and relevance of a test are established for a particular purpose (Balls et al., 1990). Over the last several decades, the European centre for validation of alternative methods (ECVAM), in co-operation with international experts, has set up guidelines for validation. Seven modules are proposed for the validity assessment of a test: test definition (scientific purpose and mechanistic basis), intra-laboratory variability, transferability, inter-laboratory variability, predictive capacity, applicability domain, and performance standard (Hartung et al., 2004).

In Europe, anyone may submit an application for validation of an alternative method to ECVAM. ECVAM then consults its network for the preliminary assessment of regulatory relevance (PARERE). PARERE then provides input, for example as to whether the novel method measures a safety endpoint for which there is a regulatory need that is currently unsatisfied. In order to prevent that the regulators in PARERE discard a proposed method for further validation because e.g. no potential use is foreseen, regular exchange between regulators and method developers on method needs and possibilities is desirable (Figure 2). If regulatory relevance is identified, ECVAM validates the method. Approval is often quicker if the applicant submits supporting evidence. In addition, validation may be expedited when the applicants have taken good *in vitro* method practice (GIVIMP) guidelines into consideration, e.g. ensured solubility of tested chemicals at the tested concentrations. ECVAM reports positive validation outcomes to its scientific advisory committee, which then performs an independent scientific review. When this step is successful a final recommendation report is submitted to the OECD, where a test guideline for the method is created. When embedded in an OECD test guideline, a novel model is then ready for regulatory use worldwide and thus ready for use in replacing animal testing. Regulatory agencies, such as ECHA and EFSA, can then easily include these methods in their guidelines (Fig 2), as has been done very quickly after the acceptance of the Extended One Generation Reproduction Toxicity (EOGRT) test, for example. These guidelines (or guidance documents) describe which safety information they require from the industry to decide whether a chemical may be allowed on the market (Figure 2), usually detailed to which methods are allowed. They are not legally binding, but are usually adhered to by the regulators and thus industry.

Validation of new *in vitro* models can also be challenging, as they cover only a small part of the body or functional system *in vivo*. This is also the case for the advanced *in vitro* systems discussed in this review, when applied to more complex safety endpoints such as repeated dose toxicity. A one-on-one comparison with the *in vivo* gold standard test to determine the predictive capacity is then not realistic. As a solution, integrated approaches to testing and assessment (IATAs) can be used, in which different alternative methods are combined to predict one endpoint. These IATAs currently form a challenge for the OECD, as multiple IATAs may be developed, consisting of different testing methods, which are not governed by a single test guideline. A new form of OECD guideline is therefore necessary to enable scientific acceptance of the advanced *in vitro* systems for more complex endpoints.

In summary, for worldwide regulatory acceptance of the currently emerging advanced *in vitro* methods:

* An exchange between regulators and developers is necessary, to exchange the regulatory needs (to avoid dismissal by PARERE) and technical possibilities;
* Test developers are strongly advised to take GIVIMP guidelines into consideration when preparing the validation package, and
* The OECD has a challenge in finding a new way to formulate mutually accepted guidelines for IATAs consisting of these methods.

In the meantime, acceptance by regulators can be achieved on a case-by-case basis for special drugs, when well justified.

# models to exemplify progress and the current state-of-the-art

## Heart

Cardiotoxicity is a major cause of drug attrition and a substantial safety concern (Cross et al., 2015; Onakpoya et al., 2016). Certain aspects of cardiotoxicity such as long QT syndrome and arrhythmia can be accurately predicted by combining hERG channel inhibition data and QTc interval measurements in the heart’s electrical cycle (Wallis, 2010). However, this leaves structural cardiotoxicity (i.e. direct damage to tissue) unaddressed. The underlying mechanisms of structural cardiotoxicity are poorly understood and current *in vitro* models cannot replicate it to an acceptable standard. Therefore, improved *in vitro* cardiotoxicitymodels are necessary.

A recurring theme between organ models is the lack of physiological relevance of 2D cultures using immortalised cell lines and dedifferentiated primary cells, which hold true for cardiotoxicity models. Furthermore, inter-species and inter-individual variation, makes extrapolation of *in vitro* data to humans challenging. Stem cell-derived cardiomyocytes (SC-CMs) are relatively novel models which help to overcome inter-individual and inter-species variation and are able to capture the phenotype of the donor cell, offering advantages over immortalised cell lines which represent only a single donor phenotype. SC-CMs have recently been shown to be accurate in predicting doxorubicin-induced cardiotoxicity severity and identifying the underlying pharmacogenetic mechanisms (Burridge et al., 2016; Mikaelian et al., 2010).

3D cardiac models show improved cell viability and enhanced structure and function (Edmondson et al., 2014; Nam et al., 2015). When developing 3D models, the origin and composition of the cells should be considered; *in vivo*, myocardial tissue comprises of 30% cardiomyocytes and 70% non-myocyte cells (NMCs, predominantly endothelial cells and fibroblasts). These NMCs are important in myocardial structure and function, as well as in development of drug induced cardiovascular injury (Brutsaert, 2003; Mikaelian et al., 2010; Souders et al., 2009). 3D models combining cardiomyocytes and NMCs were shown to be functionally superior to 2D models and could model calcium dyshomeostasis, mitochondrial disruption and loss of cell viability in response to cardiotoxicants (Pointon et al., 2013; S. Ravenscroft et al., 2016; S. M. Ravenscroft et al., 2016). Microfluidic models of cardiotoxicity are under development, they hold promise to improve physiological relevance by modelling vascularisation and structure which is not currently achievable in other models (Bhatia and Ingber, 2014).Recently, the first 3D-printed heart-on-a-chip with an integrated sensing system for non-invasive electronic readouts was produced and successfully applied to study drug responses (Lind et al., 2016).

## Skin

The skin is a highly immunocompetent barrier and is important with regard to the absorption of drugs and chemicals and therefore dermal toxicity assessment. There are several approaches towards modelling the complexity of human skin. The most common model as a simple 2D monolayers of human keratinocytes that are routinely utilised for preclinical screening. 2D monolayer models do not recapitulate aspects of skin structure such as cornification and cannot model barrier function or immunological pathways. Complex models such as reconstructed human epidermis (RHE), a 3D organotypic model formed from primary cells are capable of forming a well-stratified epithelium which can model metabolism and barrier functions (Alépée et al., 2015). There are several commercially available RHE models which have been validated by EVCAM as an alternative to animal testing for assessing skin corrosion and skin irritation in a regulatory context, while fulfilling the current OECD test guidelines (OECD, 2015, 2014). Several studies have shown the RHE model to be superior to traditional 2D culture in terms of identification of allergic sensitisers (Gibbs et al., 2013) and modelling of immunological events in the epidermal layer (Ezendam et al., 2016). However, human skin is comprised of several layers, and this complexity is not addressed by RHE. Therefore, novel models which can recapitulate this complexity are required.

In line with other multi-cell models, full thickness models (FTMs) are composed of an epidermal and dermal layer. The epidermal layer is comparable to RHE and the dermal layer contains human dermal fibroblasts distributed throughout a collagen matrix. Studies have investigated the inclusion of melanocytes and MUTZ-Langerhans cells in a FTM showing improved modelling of sensitisation (Kosten et al., 2015) and also the inclusion of other cell types including endothelial cells (Tremblay et al., 2005) and hair follicles (Michel et al., 1999). In addition to multiple cell types, the extent of cell differentiation and the formation of other skin components such as the basement membrane is also important. Creating skin models with a continuous competent stratum corneum suitable for long-term culture is challenging, but would be useful for testing of subchronic and chronic toxicity. Lab-on-chip microsystems exhibit prolonged culture times as well as stable cell functions within a controlled microenvironment and a linkage with other organs (Maschmeyer et al., 2015). However, their predictivity and validation is yet to be assessed and should be the focus of further investigations.

## Liver

Hepatotoxicity is an important safety concern for industry. Xenobiotic metabolism enzymes are highly expressed in liver to deal with the first-pass dose of drugs or chemicals. This may produce chemically reactive metabolites in addition to the parent compound, both of which may cause toxicity, making hepatotoxicity complex to accurately model. Therefore, metabolic capacity is a crucial requirement for a valid model of hepatotoxicity and this is currently not fulfilled. This is exemplified by the fact that liver toxicity is among the leading causes of safety failures in clinical trials (Cook et al., 2014; Laverty et al., 2011; Onakpoya et al., 2016). A reliable *in vitro* hepatotoxicity model should resemble the *in vivo* phenotype as well as be suitable for long term studies and high throughput screening applications (Lauschke et al., 2016).

Human primary hepatocytes (hPH), are the gold standard of *in vitro* liver models but are hindered in several respects: scarcity, since hPH may only be derived from surgical waste tissue or cadaveric liver; inter-individual variability, where the limited number of cells available from each donor and their finite lifespan can hinder repeat studies; and dedifferentiation, the loss of mature hepatic phenotype which occurs under standard cell culture conditions. Moreover, hPH cannot easily be expanded *in vitro*. Liver-derived cell lines such as HepG2 and HepaRG cells are available but have limited metabolic capacity (Sison-Young et al., 2015). Recently, proteomic characterisation of hPHs, HepG2, HepaRG and Upcytes revealed that HepaRG cells have the most similar protein profile to primary human hepatocytes (Sison-Young et al., 2015). However, in an assessment of the predictive capacity of hepatotoxicity, HepG2 was the most predictive, after the gold standard hPH (Sison-Young et al., 2016). These results indicate that none of the most commonly used hepatic models are satisfactory to fully model hepatotoxicity *in vitro*. Stem cell derived hepatocyte-like cells (SC-HLCs) are another emerging model but currently show an immature hepatic phenotype. SC-HLCs retain the phenotype of the donor cell when reprogrammed presenting the opportunity to model rare phenotypes reproducibly, which may be more susceptible to DILI

3D culture techniques such as spheroid culture (Fig 3A) often show an enhanced phenotype with superior sensitivity and specificity in detecting DILI compounds over 2D systems (Bell et al., 2016; Berger et al., 2016; Kostadinova et al., 2013; Nguyen et al., 2016). However, similar caveats to the use of hPH in 2D models apply to 3D models and the use of an expandable cell line (e.g. SC-HLCs or immortalised cell lines) may be preferred in some cases.. Spheroids form spontaneously under appropriate conditions and can be made from most extant hepatic models. It is likely that they will form a cornerstone of future hepatotoxicity testing.

In addition to 3D culture, the inclusion of non-parenchymal cells (NPCs) is necessary for modelling of DILI; inclusion of NPCs will allow modelling of complex liver toxicology such as fibrosis. NPCs can be incorporated into spheroid models (Bell et al., 2016). In addition, 3D bioprinting can reproduce the highly-organised liver architecture without use of a scaffold, with rapidly-forming tight junctions and extracellular matrix yielding solid microtissues that can model complex liver toxicity including fibrosis (Nguyen et al., 2016; Norona et al., 2016). Similarly to the other organs described herein, microfluidic devices are under development and these may prove a fruitful avenue of research in future.

## Mathematical contribution to the development of innovative toxicity models

Mathematical modelling can provide an alternative and complementary safety assessment platform for increasing mechanistic understanding, testing hypotheses *in silico*, predicting quantitative outcomes and contributing to the optimisation, design and interpretation of innovative *in vitro* models. Indeed, such quantitative approaches have always informed and influenced pharmacology, formulating models at the interface of human physiology and the drug chemistry (Kenakin and Christopoulos, 2011). The utility of quantitative modelling in drug classification and characterisation makes mathematics an important tool in drug discovery and is seen as an integral component to the advancement of toxicity testing (Krewski et al., 2007; Raies and Bajic, 2016). The computational implementation of these quantitative, mathematical models, or *in silico* modelling, is now also used extensively as part of the preclinical drug development process (Visser et al., 2014) (Fig 1). Mathematical models have been used to identify critical parameters involved in the metabolism of drugs such as acetaminophen using numerical, sensitivity and timescale analysis as well identifying critical dose thresholds (Reddyhoff et al., 2015). This illustrates how modelling can be used to optimise *in vitro* design to ensure that relevant dosing regimens are studied and that the experimental model being used is physiologically relevant in terms of the expression of appropriate metabolising enzymes for example. More directly, mathematical modelling has been used to optimise the design and operation of *in vitro* liver models such as hollow fibre membrane bioreactors by providing spatial information describing fluid and mass transport coupling in the cellular environment that would otherwise be experimentally expensive (Williams et al., 2013).

# Common Themes between Models

There common themes between advanced *in vitro* models, which may help facilitate their scientific acceptance and use. An important rule for these models is that they should be as complex as is required but as simple as possible. For example, a cell model should include the minimum number of cell types to recapitulate the *in vivo* physiology accurately. Adding more cell types can decrease reproducibility, increase costs and test failures, and lead to unnecessary interference with the test outcome. However, inclusion of multiple cell types, or features such as peristaltic movement of the tissue, can affect the tissue response to toxicants (Kim et al., 2012), presumably improving the accuracy of the prediction. The degree of accuracy required by a test, and how complex the system should be, depends on the question and on the phase within the risk assessment or drug development process. In early phases, decreased accuracy in favour of reduced costs are acceptable and simpler models may be suitable.

## Phenotyping

Determining which *in vitro* model to use for a question, particularly determining the appropriate complexity can be challenging; and proper phenotyping of advanced models can help make an informed decision. Such phenotyping of advanced models has been performed for advanced *in vitro* models of the liver, skin and heart (Berger et al., 2016; Fentem et al., 1998; Kostadinova et al., 2013; Mathur et al., 2015; Nguyen et al., 2016; Pointon et al., 2013; Reijnders et al., 2015; Wiegand et al., 2014).

The IMI MIP-DILI consortium provides a recent example of detailed phenotyping. In this study, the proteomes of hPHs cultured in 2D and 3D spheroid culture were analysed. It was shown that 2D cultures developed a dramatically altered protein expression pattern within 16h (Bell et al., 2016; Lauschke et al., 2016), while 3D spheroid cultures maintained an expression profile similar to hPH for 7 days (Bell et al., 2016). This finding indicates the greater physiological relevance of 3D cultures over 2D and suggests an improved ability to correctly identify toxicants early post-isolation of primary cells.

## Scaffolds in 3D culture

Another common theme is the use of scaffoldsto support 3D tissue structures in advanced *in vitro* models (Fig 3). Hydrogels are a type of scaffold comprised of water, extracellular matrix (ECM) proteins and growth factors that mimic the *in vivo* ECM. There are various hydrogel products available, both natural (e.g. collagen hydrogels, Matrigel, alginate) and synthetic (e.g. PuraMatrix). The natural hydrogels, notably the popular Matrigel, suffer from batch-to-batch variation. Synthetic hydrogels are well defined, have small lot-to-lot variability and can be adapted to direct functionality.

Scaffolds may also take the form of solid materials that provide a mechanical support for tissues, fabricated either from biological (e.g. collagen, fibrin, chitosan, agarose) or from synthetic (e.g. polystyrene, polycaprolactone, polyurethane) materials. The main advantage of the biological materials is their biocompatibility and flexibility, while the advantages of the synthetic materials are their consistency and controllability. Synthetic scaffolds can be fabricated with defined porosity, thickness and rigidity for the intended tissue and test. Matrix stiffness can be sensed by the cells and this can affect the types of cell adhesions, cytoskeletal structure, cell proliferation, and other factors (Yamada and Cukierman, 2007). Disadvantages common to all scaffolds include the challenge to visualise and analyse individual cells and the fact that they are not yet suitable for all cells.

Decellularised scaffolds are emerging biological scaffolds produced by removing cells from *ex vivo* tissues. Retention of the microvasculature in the ECM is an important advantage of decellularised scaffolds, allowing for easier emulation of vascular tissue. The addition of a vacuum to conventional decellularisation protocols based on enzymes or detergent, significantly reduces production time during decellularisation (Lange et al., 2015). Since availability of human tissue is limited, pig tissue is often used (Greco et al., 2015). The required tissue is taken from the pig and repopulated with human cells (e.g. *ex vivo* skin tissue to make 3D *in vitro* skin tissue). In recellularised tissue, cartilage is replaced by weaker collagen, which may lead to issues with structural integrity, furthermore, compatibility with human cells needs to be fully assessed, especially for the effects of molecular triggers that may differ in donor animals.

### Scaffold-free 3D models

Spheroids are an example of a scaffold-free model that forms spontaneously when adherent cells are denied an attachment surface (Fig 3). They are capable of producing an ECM when grown in hanging drops, ultra-low attachment conditions or in micropatterned plates (Bell et al., 2016; Kelm et al., 2003; March et al., 2015; Otsuka et al., 2004). Spheroids have an optimal size, maximally ~200 µm in diameter, to prevent necrosis in the centre due to diffusion limitations of nutrients and oxygen as these structures are not vascularised and are most often not subjected to flow. A scaffold-free system does not have the disadvantages of scaffolds, such as binding of test substances, batch-to-batch variability, and impedance of transport or diffusion from the scaffold. However, fluid flow and shear stress become difficult to incorporate in cultures lacking an anchored scaffold.

## Good Cell Culture Practice

The general considerations of Good Cell Culture Practice (GCCP) (Coecke et al., 2007) and GIVIMP are pivotal to the advance of *in vitro* method systems. This includes genotyping of cells used in culture, ensuring sufficient solubility of test substances in the medium, and correcting for loss of test substance due to evaporation, protein binding, binding to plastic, etc. In many organ-on-a-chip models, polydimethylsiloxane (PDMS) is used as material for the microfluidic chip, even though hydrophobic chemicals are known to partition into this material quite extensively (Borysiak et al., 2013; Domansky et al., 2013). This material property not only leads to loss of test substance, but also to carry-over to the medium in a next test in the same chip.

## Other Considerations

Finally, it should be stressed that the added valueof a new advanced *in vitro* model should be shown, by comparing its outcomes to those of simpler extant models. If added value is shown, ideally a comparison with a ‘gold standard’ should be made, deriving the sensitivity and specificity of the new advanced *in vitro* model. The 3D culture models discussed herein (heart, skin and liver) have not yet been fully investigated in terms of their sensitivity and specificity at predicting adverse drug reactions. Such experiments typically use a panel of training compounds such as those used by the IMI-MIP-DILI consortium where there is some understanding of the toxicological profiles of the compounds (Richert et al., 2016). By testing with a panel of training compounds at defined concentrations, novel models can be ranked in terms of sensitivity and specificity where an ideal model would be both highly sensitive to adverse drug reactions whilst maintaining high specificity. Note that for drugs, this is easily achieved, as human data are available. For other chemicals, there may only be animal data, which may be less relevant. As stated before, it should be considered that comparing the results of one organ with those of a whole organism is not a fair comparison.

Developing a novel and ideal model requires effort, creativity and innovation. Hence, collaboration between academia, regulators and industry is necessary to develop such models for the mutual benefit of all. A collaborative approach to novel model development could lead to an open access database of protocols and methodologies for developing such models, which could also serve as reference for validation and standardisation of the created models (Alépée et al., 2014).

# Conclusions and Recommendations

In conclusion, there is clearly a need for improved *in vitro* methods for toxicity testing for novel compounds such as those discussed herein. These first results with 3D models for skin, liver and heart already demonstrate added value over traditional 2D toxicity testing methods. Nevertheless, it remains to be determined how complex the models and methods should be for each (regulatory or industry) question retaining simplicity where possible. More and better characterisation of the present models, in different variants, and comparisons with less or more complex models are necessary to fully understand where they may best be deployed. This requires cross-talk between regulators, industry and developers on what the questions are ("needs"), and what added value the different variants of models may have ("possibilities"). This calls for a large research effort and for meetings and workshops to exchange the needs with the possibilities.

Finally, new methods for characterisation of novel models should be standardised, comparison against a gold standard is not always possible or appropriate for novel models and their proper handling can be challenging. Standardisation between groups for characterisation of comparable models should allow inter-model comparison and aid identification of true advances in the field.

# References

Alépée, N., Bahinski, A., Daneshian, M., De Wever, B., Fritsche, E., Goldberg, A., Hansmann, J., Hartung, T., Haycock, J., Hogberg, H.T., Hoelting, L., Kelm, J.M., Kadereit, S., McVey, E., Landsiedel, R., Leist, M., L??bberstedt, M., Noor, F., Pellevoisin, C., Petersohn, D., Pfannenbecker, U., Reisinger, K., Ramirez, T., Rothen-Rutishauser, B., Sch??fer-Korting, M., Zeilinger, K., Zurich, M.G., 2014. State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. ALTEX 31, 441–477. doi:10.14573/altex1406111

Alépée, N., Grandidier, M.H., Tornier, C., Cotovio, J., 2015. An integrated testing strategy for in vitro skin corrosion and irritation assessment using SkinEthicTM Reconstructed Human Epidermis. Toxicol. Vitr. 29, 1779–1792. doi:10.1016/j.tiv.2015.07.012

Arrowsmith, J., Miller, P., 2013. Trial Watch: Phase II and Phase III attrition rates 2011–2012. Nat. Rev. Drug Discov. 12, 569–569. doi:10.1038/nrd4090

Balls, M., Blaauboer, B., Brusick, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stammati, A.L., Walum, E., 1990. Report and Recommendations of the Caat Ergatt Workshop on the Validation of Toxicity Test Procedures. Atla-Alternatives to Lab. Anim. 18, 313–337.

Bell, C.C., Hendriks, D.F.G., Moro, S.M.L., Ellis, E., Walsh, J., Renblom, A., Fredriksson Puigvert, L., Dankers, A.C.A., Jacobs, F., Snoeys, J., Sison-Young, R.L., Jenkins, R.E., Nordling, Å., Mkrtchian, S., Park, B.K., Kitteringham, N.R., Goldring, C.E.P., Lauschke, V.M., Ingelman-Sundberg, M., 2016. Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease. Sci. Rep. 6, 25187. doi:10.1038/srep25187

Berger, B., Donzelli, M., Maseneni, S., Boess, F., Roth, A., Krähenbühl, S., Haschke, M., 2016. Comparison of Liver Cell Models Using the Basel Phenotyping Cocktail. Front. Pharmacol. 7, 1–12. doi:10.3389/fphar.2016.00443

Bhatia, S.N., Ingber, D.E., 2014. Microfluidic organs-on-chips. Nat Biotech 32, 760–772. doi:10.1038/nbt.2989 [doi]

Bode, G., Clausing, P., Gervais, F., Loegsted, J., Luft, J., Nogues, V., Sims, J., 2010. The utility of the minipig as an animal model in regulatory toxicology. J. Pharmacol. Toxicol. Methods 62, 196–220. doi:10.1016/j.vascn.2010.05.009

Borysiak, M.D., Bielawski, K.S., Sniadecki, N.J., Jenkel, C.F., Vogt, B.D., Posner, J.D., 2013. Simple replica micromolding of biocompatible styrenic elastomers. Lab Chip 13, 2773–84. doi:10.1039/c3lc50426c

Bridgland-Taylor, M.H., Hargreaves, A.C., Easter, A., Orme, A., Henthorn, D.C., Ding, M., Davis, A.M., Small, B.G., Heapy, C.G., Abi-Gerges, N., Persson, F., Jacobson, I., Sullivan, M., Albertson, N., Hammond, T.G., Sullivan, E., Valentin, J.-P., Pollard, C.E., 2006. Optimisation and validation of a medium-throughput electrophysiology-based hERG assay using IonWorksTM HT. J. Pharmacol. Toxicol. Methods 54, 189–199. doi:10.1016/j.vascn.2006.02.003

Brutsaert, D.L., 2003. Cardiac endothelial-myocardial signaling: its role in cardiac growth, contractile performance, and rhythmicity. Physiol. Rev. 83, 59–115. doi:10.1152/physrev.00017.2002

Burridge, P.W., Li, Y.F., Matsa, E., Wu, H., Ong, S.-G., Sharma, A., Holmström, A., Chang, A.C., Coronado, M.J., Ebert, A.D., Knowles, J.W., Telli, M.L., Witteles, R.M., Blau, H.M., Bernstein, D., Altman, R.B., Wu, J.C., 2016. Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity.-supple. Nat. Med. 22, 547–56. doi:10.1038/nm.4087

Coecke, S., Balls, M., Bowe, G., Davis, J., Gstraunthaler, G., Hartung, T., Hay, R., Price, A., Merten, O.W.O.-W., Stokes, W., Schechtman, L., Stacey, G., 2007. Guidance on Good Cell Culture Practice, Cell Technology for Cell Products. Springer Netherlands, Dordrecht. doi:10.1007/978-1-4020-5476-1\_49

Colatsky, T., Fermini, B., Gintant, G., Pierson, J.B., Sager, P., Sekino, Y., Strauss, D.G., Stockbridge, N., 2016. The Comprehensive in Vitro Proarrhythmia Assay (CiPA) initiative — Update on progress. J. Pharmacol. Toxicol. Methods 81, 15–20. doi:10.1016/j.vascn.2016.06.002

Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterthwaite, G., Pangalos, M.N., 2014. Lessons learned from the fate of AstraZeneca’s drug pipeline: a five-dimensional framework. Nat. Rev. Drug Discov. 13, 419–31. doi:10.1038/nrd4309

Cross, M.J., Berridge, B.R., Clements, P.J.M., Cove-Smith, L., Force, T.L., Hoffmann, P., Holbrook, M., Lyon, A.R., Mellor, H.R., Norris, A.A., Pirmohamed, M., Tugwood, J.D., Sidaway, J.E., Park, B.K., 2015. Physiological, pharmacological and toxicological considerations of drug-induced structural cardiac injury. Br. J. Pharmacol. 172, 957–974. doi:10.1111/bph.12979

Domansky, K., Leslie, D.C., McKinney, J., Fraser, J.P., Sliz, J.D., Hamkins-Indik, T., Hamilton, G.A., Bahinski, A., Ingber, D.E., 2013. Clear castable polyurethane elastomer for fabrication of microfluidic devices. Lab Chip 13, 3956. doi:10.1039/c3lc50558h

Edmondson, R., Broglie, J.J., Adcock, A.F., Yang, L., 2014. Three-Dimensional Cell Culture Systems and Their Applications in Drug Discovery and Cell-Based Biosensors. Assay Drug Dev. Technol. 12, 207–218. doi:10.1089/adt.2014.573

Eichler, H.-G., Pignatti, F., Flamion, B., Leufkens, H., Breckenridge, A., 2008. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. Nat. Rev. Drug Discov. 7, 818–826. doi:10.1038/nrd2664

Ezendam, J., Braakhuis, H.M., Vandebriel, R.J., 2016. State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies. Arch. Toxicol. 90, 2861–2883. doi:10.1007/s00204-016-1842-4

Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhütter, H.G., Liebsch, M., 1998. The ECVAM international validation study on in vitro tests for skin corrosivity. 2. Results and evaluation by the management team. Toxicol. Vitr. 12, 483–524. doi:10.1016/S0887-2333(98)00019-8

Finn, M.A., Stark, J.F., 2015. Medical science and the Cruelty to Animals Act 1876: A re-examination of anti-vivisectionism in provincial Britain. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. 49, 12–23. doi:10.1016/j.shpsc.2014.10.007

Gibbs, S., Corsini, E., Spiekstra, S.W., Galbiati, V., Fuchs, H.W., DeGeorge, G., Troese, M., Hayden, P., Deng, W., Roggen, E., 2013. An epidermal equivalent assay for identification and ranking potency of contact sensitizers. Toxicol. Appl. Pharmacol. 272, 529–541. doi:10.1016/j.taap.2013.07.003

Greaves, P., Williams, A., Eve, M., 2004. First dose of potential new medicines to humans: how animals help. Nat. Rev. Drug Discov. 3, 226–236. doi:10.1038/nrd1329

Greco, K. V, Francis, L., Somasundaram, M., Greco, G., English, N.R., Roether, J.A., Boccaccini, A.R., Sibbons, P., Ansari, T., 2015. Characterisation of porcine dermis scaffolds decellularised using a novel non-enzymatic method for biomedical applications. J. Biomater. Appl. 30, 239–53. doi:10.1177/0885328215578638

Hartung, T., 2010. Food for thought...on alternative methods for chemical safety testing. ALTEX 27, 3–14.

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V., 2004. A modular approach to the ECVAM principles on test validity. ATLA Altern. to Lab. Anim.

Heringa, M.B., de Wit, L., Bos, P.M.J., Hakkert, B., 2014. Do current EU regulations for the safety assessment of chemical substances pose legal barriers for the use of alternatives to animal testing?, RIVM Report. Bilthoven.

Kelm, J.M., Timmins, N.E., Brown, C.J., Fussenegger, M., Nielsen, L.K., 2003. Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types. Biotechnol. Bioeng. 83, 173–80. doi:10.1002/bit.10655

Kemsley, J., 2015. How The Internet Enabled Information Sharing Through Massive Databases. Chem. Eng. News 93, 14–15.

Kenakin, T., Christopoulos, A., 2011. Analytical pharmacology: The impact of numbers on pharmacology. Trends Pharmacol. Sci. 32, 189–196. doi:10.1016/j.tips.2011.01.002

Kenny, P.A., Lee, G.Y., Myers, C.A., Neve, R.M., Semeiks, J.R., Spellman, P.T., Lorenz, K., Lee, E.H., Barcellos-Hoff, M.H., Petersen, O.W., Gray, J.W., Bissell, M.J., 2007. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Mol. Oncol. 1, 84–96. doi:10.1016/j.molonc.2007.02.004

Kessel, M., Frank, F., 2007. A better prescription for drug-development financing. Nat. Biotechnol. 25, 859–866. doi:10.1038/nbt0807-859

Kim, H.J., Huh, D., Hamilton, G., Ingber, D.E., 2012. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. Lab Chip 12, 2165. doi:10.1039/c2lc40074j

Kirkland, D., Zeiger, E., Madia, F., Corvi, R., 2014. Mutation Research / Genetic Toxicology and Environmental Mutagenesis Can in vitro mammalian cell genotoxicity test results be used to complement positive results in the Ames test and help predict carcinogenic or in vivo genotoxic activity ? II . Construct. Mutat. Res. - Genet. Toxicol. Environ. Mutagen. 775–776, 69–80. doi:10.1016/j.mrgentox.2014.10.006

Kostadinova, R., Boess, F., Applegate, D., Suter, L., Weiser, T., Singer, T., Naughton, B., Roth, A., 2013. A long-term three dimensional liver co-culture system for improved prediction of clinically relevant drug-induced hepatotoxicity. Toxicol. Appl. Pharmacol. 268, 1–16. doi:10.1016/j.taap.2013.01.012

Kosten, I.J., Spiekstra, S.W., de Gruijl, T.D., Gibbs, S., 2015. MUTZ-3 derived Langerhans cells in human skin equivalents show differential migration and phenotypic plasticity after allergen or irritant exposure. Toxicol. Appl. Pharmacol. 287, 35–42. doi:10.1016/j.taap.2015.05.017

Krewski, D., Acosta, D.J., Andersen, M., Bailar III, J., Boekelhedie, K., Brent, R., Charneley, G., Cheung, V., Green, S., Kelsey, K., Kerkvliet, N., Li, A., McCray, L., Meyer, O., Patterson, R.D., Pennie, W., Scala, R., Solomon, G., Stephens, M., Yager, J., Zeise, L., 2007. Toxicity Testing in the 21St Century: a Vision and a Strategy. Natl. Acad. Sci. 13, 51–138. doi:10.1080/10937404.2010.483176.TOXICITY

Lange, P., Greco, K., Partington, L., Carvalho, C., Oliani, S., Birchall, M.A., Sibbons, P.D., Lowdell, M.W., Ansari, T., 2015. Pilot study of a novel vacuum-assisted method for decellularization of tracheae for clinical tissue engineering applications. J. Tissue Eng. Regen. Med. n/a-n/a. doi:10.1002/term.1979

Lauschke, V.M., Hendriks, D.F.G., Bell, C.C., Andersson, T.B., Ingelman-Sundberg, M., 2016. Novel 3D Culture Systems for Studies of Human Liver Function and Assessments of the Hepatotoxicity of Drugs and Drug Candidates. Chem. Res. Toxicol. acs.chemrestox.6b00150. doi:10.1021/acs.chemrestox.6b00150

Laverty, H.G., Benson, C., Cartwright, E.J., Cross, M.J.M., Garland, C., Hammond, T., Holloway, C., McMahon, N., Milligan, J., Park, B.K.B., Pirmohamed, M., Pollard, C., Radford, J., Roome, N., Sager, P., Singh, S., Suter, T., Suter, W., Trafford, A., Volders, P.G.A., Wallis, R., Weaver, R., York, M., Valentin, J.P.J., 2011. How can we improve our understanding of cardiovascular safety liabilities to develop safer medicines? Br. J. Pharmacol. 163, 675–693. doi:10.1111/j.1476-5381.2011.01255.x

Lind, J.U., Busbee, T.A., Valentine, A.D., Pasqualini, F.S., Yuan, H., Yadid, M., Park, S., Kotikian, A., Nesmith, A.P., Campbell, P.H., Vlassak, J.J., Lewis, J.A., Parker, K.K., 2016. Instrumented cardiac microphysiological devices via multimaterial three-dimensional printing. Nat. Mater. 1. doi:10.1038/NMAT4782

March, S., Ramanan, V., Trehan, K., Ng, S., Galstian, A., Gural, N., Scull, M. a, Shlomai, A., Mota, M.M., Fleming, H.E., Khetani, S.R., Rice, C.M., Bhatia, S.N., 2015. Micropatterned coculture of primary human hepatocytes and supportive cells for the study of hepatotropic pathogens. Nat. Protoc. 10, 2027–53. doi:10.1038/nprot.2015.128

Maschmeyer, I., Lorenz, A.K., Schimek, K., Hasenberg, T., Ramme, A.P., Hübner, J., Lindner, M., Drewell, C., Bauer, S., Thomas, A., Sambo, N.S., Sonntag, F., Lauster, R., Marx, U., 2015. A four-organ-chip for interconnected long-term co-culture of human intestine, liver, skin and kidney equivalents. Lab Chip 15, 2688–2699. doi:10.1039/C5LC00392J

Mathes, S.H., Ruffner, H., Graf-Hausner, U., 2014. The use of skin models in drug development. Adv. Drug Deliv. Rev. 69–70, 81–102. doi:10.1016/j.addr.2013.12.006

Mathur, A., Loskill, P., Shao, K., Huebsch, N., Hong, S., Marcus, S.G., Marks, N., Mandegar, M., Conklin, B.R., Lee, L.P., Healy, K.E., 2015. Human iPSC-based cardiac microphysiological system for drug screening applications. Sci. Rep. 5, 8883. doi:10.1038/srep08883

McKenzie, R., Fried, M.W., Sallie, R., Conjeevaram, H., Di Bisceglie, A.M., Park, Y., Savarese, B., Kleiner, D., Tsokos, M., Luciano, C., 1995. Hepatic failure and lactic acidosis due to fialuridine (FIAU), an investigational nucleoside analogue for chronic hepatitis B. N. Engl. J. Med. 333, 1099–1105. doi:10.1056/NEJM199510263331702

Megit, S., 2011. Immunocore pioneers new safety studies. MedNous 5, 14–15.

Michel, M., L’Heureux, N., Pouliot, R., Xu, W., Auger, F. a, Germain, L., 1999. Characterization of a new tissue-engineered human skin equivalent with hair. In Vitro Cell. Dev. Biol. Anim. 35, 318–326. doi:10.1007/s11626-999-0081-x

Mikaelian, I., Buness, A., de Vera-Mudry, M.-C., Kanwal, C., Coluccio, D., Rasmussen, E., Char, H.W., Carvajal, V., Hilton, H., Funk, J., Hoflack, J.-C., Fielden, M., Herting, F., Dunn, M., Suter-Dick, L., 2010. Primary Endothelial Damage Is the Mechanism of Cardiotoxicity of Tubulin-Binding Drugs. Toxicol. Sci. 117, 144–151. doi:10.1093/toxsci/kfq189

Moeller, T.A., Shukla, S.J., Xia, M., 2012. Assessment of Compound Hepatotoxicity Using Human Plateable Cryopreserved Hepatocytes in a 1536-Well-Plate Format. Assay Drug Dev. Technol. 10, 78–87. doi:10.1089/adt.2010.0365

Moore, N., 2016. Lessons from the fatal French study BIA-10-2474. BMJ 353, i2727. doi:10.1136/bmj.i2727

Nam, K.-H., Smith, A.S.T., Lone, S., Kwon, S., Kim, D.-H., 2015. Biomimetic 3D Tissue Models for Advanced High-Throughput Drug Screening. J. Lab. Autom. 20, 201–15. doi:10.1177/2211068214557813

Netherlands National Committee for the protection of animals used for scientific purposes, 2016. Transition to non-animal research.

Nguyen, D.G., Funk, J., Robbins, J.B., Crogan-Grundy, C., Presnell, S.C., Singer, T., Roth, A.B., 2016. Bioprinted 3D Primary Liver Tissues Allow Assessment of Organ-Level Response to Clinical Drug Induced Toxicity In Vitro. PLoS One 11, e0158674. doi:10.1371/journal.pone.0158674

Norona, L.M., Nguyen, D.G., Gerber, D.A., Presnell, S.C., LeCluyse, E.L., 2016. Modeling Compound-Induced Fibrogenesis *In Vitro* Using Three-Dimensional Bioprinted Human Liver Tissues. Toxicol. Sci. kfw169. doi:10.1093/toxsci/kfw169

OECD, 2015. Test No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. doi:http://dx.doi.org/10.1787/9789264242845-en

OECD, 2014. Test No. 431: In Vitro Skin Corrosion: Human Skin Model Test. doi:http://dx.doi.org/10.1787/9789264071148-en

Onakpoya, I.J., Heneghan, C.J., Aronson, J.K., 2016. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. BMC Med. 14, 10. doi:10.1186/s12916-016-0553-2

Otsuka, H., Hirano, A., Nagasaki, Y., Okano, T., Horiike, Y., Kataoka, K., 2004. Two-dimensional multiarray formation of hepatocyte spheroids on a microfabricated PEG-brush surface. ChemBioChem 5, 850–855. doi:10.1002/cbic.200300822

Pignatti, F., Gravanis, I., Herold, R., Vamvakas, S., Jonsson, B., Marty, M., 2011. The European Medicines Agency: An Overview of Its Mission, Responsibilities, and Recent Initiatives in Cancer Drug Regulation. Clin. Cancer Res. 17, 5220–5225. doi:10.1158/1078-0432.CCR-11-0623

Pointon, A., Abi-gerges, N., Cross, M.J., Sidaway, J.E., 2013. Phenotypic profiling of structural cardiotoxins in vitro reveals dependency on multiple mechanisms of toxicity. Toxicol. Sci. 132, 317–326. doi:10.1093/toxsci/kft005

Raies, A.B., Bajic, V.B., 2016. In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdiscip. Rev. Comput. Mol. Sci. 6, 147–172. doi:10.1002/wcms.1240

Ravenscroft, S., Bauch, C., Hinton, L., Walker, P., 2016. Microtissues for in vitro toxicity assessment: Cost effective and in vivo relevant toxicology tools. Drug Discov. World.

Ravenscroft, S.M., Pointon, A., Williams, A.W., Cross, M.J., Sidaway, J.E., 2016. Cardiac non-myocyte cells show enhanced pharmacological function suggestive of contractile maturity in stem cell derived cardiomyocyte microtissues. Toxicol. Sci. 152, 99–112. doi:10.1093/toxsci/kfw069

Reddyhoff, D., Ward, J., Williams, D., Regan, S., Webb, S., 2015. Timescale analysis of a mathematical model of acetaminophen metabolism and toxicity. J. Theor. Biol. 386, 132–146. doi:10.1016/j.jtbi.2015.08.021

Reijnders, C.M.A., van Lier, A., Roffel, S., Kramer, D., Scheper, R.J., Gibbs, S., 2015. Development of a Full-Thickness Human Skin Equivalent In Vitro Model Derived from TERT-Immortalized Keratinocytes and Fibroblasts. Tissue Eng. Part A 21, 2448–59. doi:10.1089/ten.TEA.2015.0139

Richert, L., Baze, A., Parmentier, C., Gerets, H.H.J., Sison-Young, R., Dorau, M., Lovatt, C., Czich, A., Goldring, C., Park, B.K., Juhila, S., Foster, A.J., Williams, D.P., 2016. Cytotoxicity evaluation using cryopreserved primary human hepatocytes in various culture formats. Toxicol. Lett. 258, 207–215. doi:10.1016/j.toxlet.2016.06.1127

Rimann, M., Graf-Hausner, U., 2012. Synthetic 3D multicellular systems for drug development. Curr. Opin. Biotechnol. 23, 803–809. doi:10.1016/j.copbio.2012.01.011

Russell, W.M.S.W.M.S., Burch, R.L.R.L., Hume, C.W., 1959. The principles of humane experimental technique, UFAW, London. Methuen London, London: Methuen & Co. Ltd.

Senderowicz, A.M., 2010. Information Needed to Conduct First-in-Human Oncology Trials in the United States: A View from a Former FDA Medical Reviewer. Clin. Cancer Res. 16, 1719–1725. doi:10.1158/1078-0432.CCR-09-2766

Shukla, S.J., Huang, R., Austin, C.P., Xia, M., 2010. The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. Drug Discov. Today 15, 997–1007. doi:10.1016/j.drudis.2010.07.007

Silbergeld, E.K., Mandrioli, D., Cranor, C.F., 2015. Regulating Chemicals: Law, Science, and the Unbearable Burdens of Regulation. Annu. Rev. Public Health 36, 175–191. doi:10.1146/annurev-publhealth-031914-122654

Sison-Young, R.L., Lauschke, V.M., Johann, E., Alexandre, E., Antherieu, S., Aerts, H., Gerets, H.H.J., Labbe, G., Ho??t, D., Dorau, M., Schofield, C.A., Lovatt, C.A., Holder, J.C., Stahl, S.H., Richert, L., Kitteringham, N.R., Jones, R.P., Elmasry, M., Weaver, R.J., Hewitt, P.G., Ingelman-Sundberg, M., Goldring, C.E., Park, B.K., 2016. A multicenter assessment of single-cell models aligned to standard measures of cell health for prediction of acute hepatotoxicity. Arch. Toxicol. 1–16. doi:10.1007/s00204-016-1745-4

Sison-Young, R.L.C., Mitsa, D., Jenkins, R.E., Mottram, D., Alexandre, E., Richert, L., Aerts, H., Weaver, R.J., Jones, R.P., Johann, E., Hewitt, P.G., Ingelman-Sundberg, M., Goldring, C.E.P., Kitteringham, N.R., Park, B.K., 2015. Comparative Proteomic Characterization of 4 Human Liver-Derived Single Cell Culture Models Reveals Significant Variation in the Capacity for Drug Disposition, Bioactivation, and Detoxication. Toxicol. Sci. 147, 412–424. doi:10.1093/toxsci/kfv136

Souders, C.A., Bowers, S.L.K., Baudino, T.A., 2009. Cardiac fibroblast: The renaissance cell. Circ. Res. doi:10.1161/CIRCRESAHA.109.209809

Spielmann, H., Grune, B., Liebsch, M., Seiler, A., Vogel, R., 2008. Successful validation of in vitro methods in toxicology by ZEBET, the National Centre for Alternatives in Germany at the BfR (Federal Institute for Risk Assessment). Exp. Toxicol. Pathol. 60, 225–233. doi:10.1016/j.etp.2008.01.012

Suntharalingam, G., Perry, M.R., Ward, S., Brett, S.J., Castello-Cortes, A., Brunner, M.D., Panoskaltsis, N., 2006. Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. N.Engl.J.Med. 355, 1018–1028. doi:10.1056/NEJMoa063842

Tremblay, P.-L., Berthod, F., Germain, L., Auger, F.A., 2005. In vitro evaluation of the angiostatic potential of drugs using an endothelialized tissue-engineered connective tissue. J. Pharmacol. Exp. Ther. 315, 510–516. doi:jpet.105.089524 [pii]\r10.1124/jpet.105.089524

Visser, S.A.G., de Alwis, D.P., Kerbusch, T., Stone, J.A., Allerheiligen, S.R.B., 2014. Implementation of Quantitative and Systems Pharmacology in Large Pharma. CPT Pharmacometrics Syst. Pharmacol. 3, e142. doi:10.1038/psp.2014.40

Vonk, R., van de Laar, C.W.E., Hegger, I., Hoebert, J.M., 2015. Legal barriers for the use of alternatives to animal testing: do current EU regulations and guidelines for regulatory acceptance of medicinal products pose legal barriers? Bilthoven. doi:10.13140/RG.2.1.1130.6727

Wallis, R.M., 2010. Integrated risk assessment and predictive value to humans of non-clinical repolarization assays. Br. J. Pharmacol. 159, 115–121. doi:10.1111/j.1476-5381.2009.00395.x

Wiegand, C., Hewitt, N.J., Merk, H.F., Reisinger, K., 2014. Dermal xenobiotic metabolism: A comparison between native human skin, four in vitro skin test systems and a liver system. Skin Pharmacol. Physiol. 27, 263–275. doi:10.1159/000358272

Williams, D.P., Shipley, R., Ellis, M.J., Webb, S., Ward, J., Gardner, I., Creton, S., 2013. Novel in vitro and mathematical models for the prediction of chemical toxicity. Toxicol. Res. 2, 40–59. doi:10.1039/C2TX20031G

Yamada, K.M., Cukierman, E., 2007. Modeling Tissue Morphogenesis and Cancer in 3D. Cell 130, 601–610. doi:10.1016/j.cell.2007.08.006

Zbinden, G., 1993. The Concept of Multispecies Testing in Industrial Toxicology. Regul. Toxicol. Pharmacol. 17, 85–94. doi:10.1006/rtph.1993.1009

**Fig 1** *Safety assessments during the development of chemicals or drugs and opportunities for the application of innovative in vitro models* a) Number of newly synthesized organic and inorganic chemical substances recorded in the CAS Registry (Kemsley, 2015) and number of substances registered at the European Chemicals Agency (ECHA). b) Schematics of the development process for chemical substances including safety assessments. Violet squares show interactions with regulatory authorities with respect to chemical safety reports (CSR) and the provision of safety data sheets (SDS). Opportunities for the application of 3D or organotypic *in vitro* models are indicated. c) Schematics of the drug development pipeline from the identification of safety liabilities during discovery, screening and early development to risk/benefit assessments during clinical trials and product life cycle management. Violet squares show important interactions with regulatory authorities, e.g. Investigational New Drug (IND) applications and New Drug or Biologic License Applications (NDA/BLA). Indicated are the major organ systems involved in pre-clinical and clinical safety failures (Cook et al., 2014) as well as opportunities for the application of 3D or organotypic *in vitro* models.

**Fig 2** *Scheme of the three main actors in the development of safety test methods for chemical substances and their mutual relations and roles in the EU.*The regulators lay down the required safety information for allowing a substance on the market, industry provides this information to the regulators, which assess it. Industry gains this information through safety tests, which are sometimes developed and performed in-house, but often obtained from test developers or commissioned to contract research organizations (CROs). Industry then sends their substances to the CRO. Industry can also fund test developers, such as universities, to develop or validate certain desired tests. Test developers and CROs can also obtain funds from regulators, who can also indicate directly (i.e. not through industry) which test methods are needed. The limitations of current tests and possibilities of new technologies need to be communicated to the regulators, so these can adjust the safety information requirements accordingly. Such information can also be provided by industry, but will then not be free of commercial interests.

Fig 3 *Approaches to enable 3D architecture for a cell culture* production of spheroids in e.g. hanging drops (A), use of hydrogels (B) of synthetic or natural materials, such as alginate (last picture in B), or use of scaffolds produced from natural materials or synthetically from polymers (C), which are either layered (first picture), electrospun (second picture) or polymerized in a sponge-like structure (third picture)