



Novel Technologies and an Overall Strategy to Allow Hazard Assessment and Risk Prediction of Chemicals, Cosmetics, and Drugs with Animal-Free Methods

Marcel Leist^{1,2}, Brett A. Lidbury³, Chihae Yang⁴, Patrick J. Hayden⁵, Jens M. Kelm⁶, Stephanie Ringeissen⁷, Ann Detroyer⁷, Jean R. Meunier⁷, James F. Rathman⁸, George R. Jackson, Jr.⁵, Gina Stolper⁵, and Nina Hasiwa¹

¹Centre for Alternatives to Animal Testing, CAAT-Europe, University of Konstanz, Germany; ²Doerenkamp-Zbinden Chair of in-vitro toxicology and biomedicine, University of Konstanz, Germany; ³MAWA Fellow, John Curtin School of Medical Research, Australian National University, Canberra, Australia; ⁴Altamira, LLC, Columbus, OH, USA; ⁵MatTek Corp., Ashland, USA; ⁶InSphero AG, Zurich, Switzerland; ⁷L'Oréal R&I, Aulnay-sous-Bois, France; ⁸The Ohio State University, Columbus, OH, USA

Summary

Several alternative methods to replace animal experiments have been accepted by legal bodies. An even larger number of tests are under development or already in use for non-regulatory applications or for the generation of information stored in proprietary knowledge bases. The next step for the use of the different in vitro methods is their combination into integrated testing strategies (ITS) to get closer to the overall goal of predictive “in vitro-based risk evaluation processes.” We introduce here a conceptual framework as the basis for future ITS and their use for risk evaluation without animal experiments. The framework allows incorporation of both individual tests and already integrated approaches. Illustrative examples for elements to be incorporated are drawn from the session “Innovative technologies” at the 8th World Congress on Alternatives and Animal Use in the Life Sciences, held in Montreal, 2011. For instance, LUHMES cells (conditionally immortalized human neurons) were presented as an example for a 2D cell system. The novel 3D platform developed by InSphero was chosen as an example for the design and use of scaffold-free, organotypic microtissues. The identification of critical pathways of toxicity (PoT) may be facilitated by approaches exemplified by the MatTek 3D model for human epithelial tissues with engineered toxicological reporter functions. The important role of in silico methods and of modeling based on various pre-existing data is demonstrated by Altamira’s comprehensive approach to predicting a molecule’s potential for skin irritancy. A final example demonstrates how natural variation in human genetics may be overcome using data analytic (pattern recognition) techniques borrowed from computer science and statistics. The overall hazard and risk assessment strategy integrating these different examples has been compiled in a graphical work flow.

Keywords: risk assessment, animal-free methods, Tox21, artificial tissue, systems toxicology

1 Introduction

Current toxicological risk assessment to ensure human health or the safety of the environment is still based predominantly on animal studies. Besides ethical considerations, there are also scientific reasons to stop the use of animals and to switch to scientifically validated modern test systems, which may provide a deeper insight into the mechanisms of toxicity.

As science and technology continuously move ahead, and new innovations appear on the market daily, we suggest taking advantage of this situation by combining good laboratory prac-

tice and solid toxicological experience with new concepts and innovative technologies (van Thriel et al., 2012).

Traditional approaches are rooted mainly in the fields of biology, medicine, and veterinary sciences (Leist et al., 2012). New technologies are emerging from many additional fields (Leist et al., 2008a,b). These include computer information sciences, statistics, mechanical engineering, molecular biology, drug discovery, and many others. Many technologies, approaches, and assays have been described individually. Recently, more integrated strategies for using combinations of *in vitro* data and *in silico* models to predict drug and chemical

Received August 12, 2012; accepted in revised form October 17, 2012



safety have emerged as well. One of the most comprehensive strategies to make toxicology largely independent of animal studies is “*Toxicity testing in the 21st century: a vision and a strategy*,” as put forward by the National Research Council (NRC, 2007). This has triggered the design of new concepts and the setup of research programs dedicated to the implementation of this vision (Hengstler et al., 2012). Some of these programs, such as ToxCast, focus mainly on high-throughput and use mostly technology and methods suitable for this purpose as building blocks (Kavlock et al., 2012; Rotroff et al., 2012). For example, a high-throughput risk assessment (HTRA) framework was published last year (Judson et al., 2011). A different approach makes use of relatively complex models that have been established primarily with the purpose of replacing defined animal experiments.

The approach taken here is to suggest a comprehensive theoretical strategy to undertake risk assessment on the basis of available alternative test systems of differing complexity and with varying throughput. This strategy may be used as a framework to look for, identify, and employ new technologies. It may also be used to incorporate such diverse technologies as high-throughput screening of simple biochemical endpoints and the evaluation of complex functional endpoints in 3D engineered tissue. The different available and newly emerging methods would then support individual steps outlined by the strategy and form the basis for translating the theoretical framework into practice. Such a risk assessment scheme is outlined here, and a heterogeneous set of new technologies was chosen as an example of how to apply the new approaches within this overall framework.

2 *In vitro* based risk evaluation approach

The development of suitable *in vitro* methods for risk assessment is an ongoing effort, and many promising methods are available already. We suggest a general scheme, which can then be filled in with suitable methods according to the individual requirements (Fig. 1). The basic idea of the scheme was also discussed in the context of biomarker usage (Blaauboer et al., 2012).

Potential starting steps for compound evaluation are based on literature review, *in silico* as well as purely physicochemical and biochemical methods to define the compound in question, its intracellular distribution, and its potential metabolites. These steps would be followed by the application of a battery of methods to define the biokinetic behavior of the compound and metabolites in *in vitro* systems and to obtain essential data for physiology-based pharmacokinetic modeling (PBPK). Some of this information would be used immediately for subsequent steps, while other information (exposure information) would be used later for the final risk assessment.

The core of the strategy is the concept that *in silico* evaluation and thorough literature research go hand in hand with an *in vitro*-based hazard assessment. These are the cornerstones of any type of “*in vitro* risk assessment.”

In principle, any point of the circle may be used as starting point, as reconsideration and re-iteration is possible and necessary to ensure that all data and aspects are taken into account. If new compounds are to be evaluated, it might make sense to start with an exposure assessment in order to avoid testing of concentrations that are unlikely to occur in real life. “*Exposure Assessment is the process of estimating or measuring the magnitude, frequency and duration of exposure to an agent, along with the number and characteristics of the population exposed. Ideally, it describes the sources, pathways, routes, and the uncertainties in the assessment*” (IPCS, 2004). Risk assessment would, for instance, be handled differently if the compound under investigation is used only as intermediate in a closed chemical production process or if it is part of a cosmetic product, or if it is an environmental chemical known to enter the food chain. Data on exposure may be scarce, of low reliability, or not available at all, especially for environmental agents. Under such circumstances, the scheme suggests moving to subsequent steps and evaluating compound hazard first. For the risk assessment, exposure would then need to be estimated roughly, classified on the basis of use categories and on the knowledge of environmental fate (e.g., transport models, persistence, bioaccumulation, and transformation). In the absence of all information, risk classification would need to be heavily based on hazard information and then adapted as more exposure information becomes available.

A thorough literature review¹ is usually the basis of a successful evaluation. Defining a search strategy is necessary to ensure coverage of the topic, as the “state-of-the-art” should be the starting point. In this sense, the term “literature” is meant to cover all kinds of databases that are accessible. The literature review would initially focus on the compound, but in further iterations it also would cover the methods used for hazard assessment, as well as other technologies used for evaluation of the compound. In addition, experiments that did not work and information on failed tests or compounds might be very helpful. Unfortunately, this information is seldom freely available.

Due to progress in science and technology, *in silico* modeling becomes more and more a prerequisite in the risk assessment field, and it was specified by Hartung and Hoffmann in 2009 as “*anything that we can do with a computer in toxicology*” (Hartung and Hoffmann, 2009). These days, it is difficult to define “*in silico* toxicology” exactly, as *in silico* components are present in practically every area of risk assessment (Raunio, 2011), and non-testing data can be generated by several approaches, including: grouping approaches, which consist of read-across and chemical category formation, structure-activity relationship (SAR) and quantitative SAR (QSAR)². A structural physicochemical reactivity characterization of a compound is currently done routinely (Valerio, 2009, 2011).

Another very important point is the determination of the biokinetic behavior of the compound in all test systems used. This comprises information on the real free concentration in

¹ <http://www.ctu.edu.vn/guidelines/scientific/scientific/2.1howliteraturesearch.htm>

² <http://echa.europa.eu/web/guest/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment>

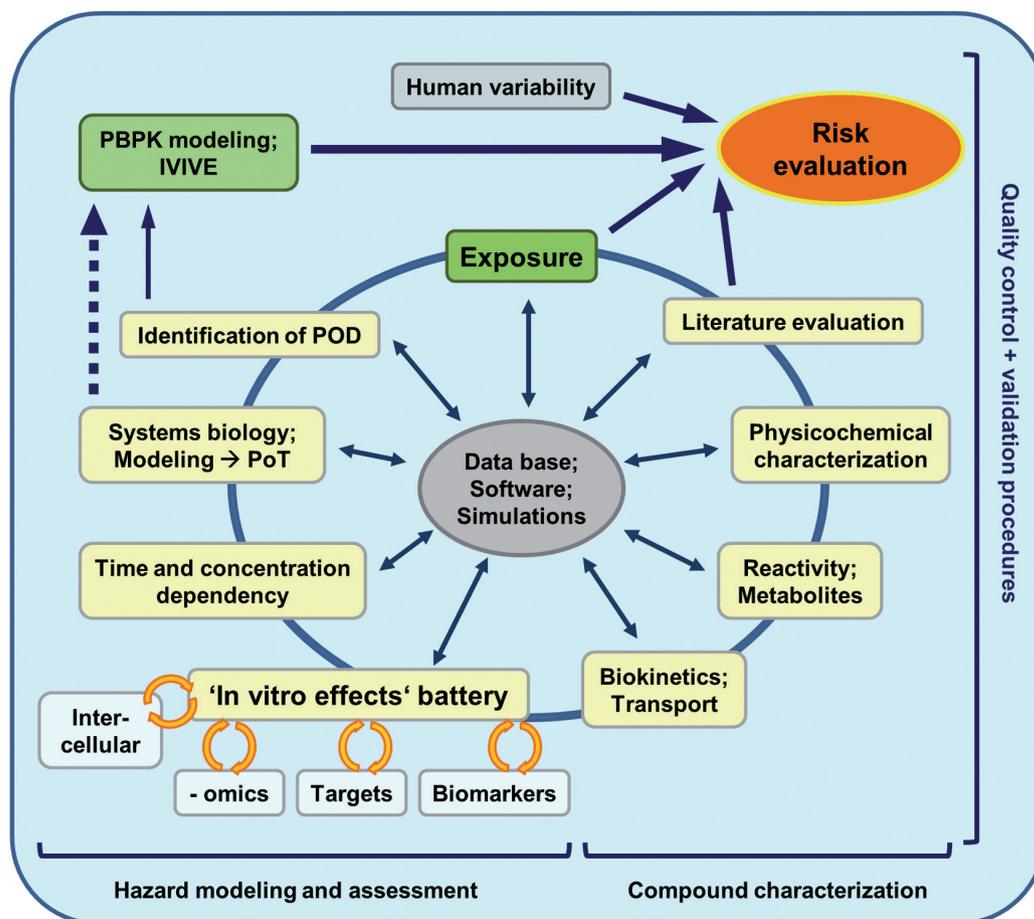


Fig. 1: *In vitro*-based risk evaluation approach

The overall strategy relies on two essential steps: First, important background information on the compound and its *in vitro* hazard is assessed (circle of yellow boxes). Then, an *in vitro-in vivo* extrapolation (IVIVE) is used to predict relevant human doses. These are then related to different exposure scenarios, taking also different human subpopulations into account to arrive finally at a risk evaluation (upper part of the scheme). The part of the strategy focusing on *in vitro* hazard assessment would start with the identification of appropriate tests and conditions, based on the information gained from the initial steps. An *in vitro* test battery with various 2D and 3D models (see sections 3-5) would be used to obtain detailed concentration-response information to arrive at a point of departure (POD) for IVIVE. In most cases, steps of the cycle may need to be revisited in an iterative process of optimization. All steps would yield information to databases, simulations, and new software solutions, and *vice versa*. Besides these general interactions, the specific hazard evaluation of a compound would involve a process of modeling and systems biology approach to use all the data generated for the identification/mapping of key pathways of toxicity (PoT) and for definition of thresholds of their activation. A combination of such steps is demonstrated in section 6. Once a defined POD (= critical concentration of a compound to trigger hazardous effects) has been derived, IVIVE will help to define the hazardous dose in humans. This may be affected, e.g., by genetic variability (as outlined in section 7). Combination of human hazard information and exposure data will finally support the risk evaluation process.

the cell culture medium, and ideally also in different cellular compartments. The latter often will require information on transport processes of the compound across different membranes. To perform an *in vitro-in vivo* extrapolation (IVIVE), the real concentration of a compound has to be determined (Coecke et al., 2012). Some compounds may bind to plastic or to proteins or they may evaporate; others react quickly or become metabolized. Therefore, the freely available concentration is not necessarily identical to the nominal concentration. In case of metabolism of the compound within the experimen-

tal system, similar information is necessary on the different metabolites. Such background information on the chemical facilitates identification of tests and conditions to set-up an *in vitro* effects battery.

The *in vitro* effects battery could include simple model organisms such as drosophila or zebra fish (Sipes et al., 2011b; Padilla et al., 2012), and it would include assays for functional changes, to provide broad omics information (e.g., metabolomics, transcriptomics), and to define biochemical targets of the chemicals. In addition to the mostly correlative methods,



this testing step also would use approaches to define key toxicity events and/or biomarkers more causally, such as reporter assays and inhibition of suspected pathways.

Such battery approaches are well known from the field of genotoxicity (Kirkland et al., 2011). They are under active development for the area of skin sensitization, and it is expected that they will be broadly applied in the future in many other domains, such as developmental neurotoxicity (Kadereit et al., 2012). The battery should be validated carefully, including appropriate controls. The tests would generate *in vitro* concentration response curves. This should give a first idea of the biological profile of a compound. These data and their use in an appropriate integrated model should yield a point of departure (POD). The POD is defined as the concentration that results in a significant hazard-related change in the *in vitro* system, which is considered predictive for the *in vivo* situation. Different systems with different endpoints, considered also at different incubation times, will result in a large set of data on the compound's effects at different concentrations. This will be particularly important, as test systems diversify. In addition to the "classical tests," mimicking certain complex modes of action (e.g., inflammatory activation of cells, or disturbed differentiation), new test systems will either test only the effects of compounds on defined biochemical targets (e.g., receptor activation or enzyme inhibition), or they will obtain broad information by different omics technologies (metabolomics, transcriptomics, proteomics). Moreover, some of the test systems will also evaluate compound effects in 3D systems and between different cell types, e.g., between microglia and astrocytes (Kuegler et al., 2012; Schildknecht et al., 2011). Some of the effects measured may not be related to toxicity but may be epiphenomena or cellular counter-regulations. Sometimes, the effects observed also may depend on the metabolic state of the cells. Modeling such different situations can be useful to predict toxicity under greatly varying conditions in humans (Latta et al., 2000; Falsig et al., 2004). In most cases, extensive modeling will be required to process the data and to determine the relevant pathways of toxicity (PoT), as well as the threshold concentrations that trigger them. These will then be used as POD for IVIVE (Blaauboer, 2010). When, finally, the variability and sensitivity of human subpopulations (Bolt et al., 2003; Pohl and Scinicariello, 2011; Mirlohi et al., 2011) are taken into account, this overall procedure may lead to a risk evaluation without the use of animals.

3 Use of engineered human cells for complex 2D models of toxicity

To start the risk assessment process, it is necessary to combine different biological methods to finally reach a prediction about the hazard of a substance. Relatively simple 2D systems may be a good starting point. Preferably, human cells should be used, but these often require technologically challenging procedures. Pluripotent stem cells may be a good general source of different cell types (Balmer et al., 2012).

Here, the LUHMES cell system is presented as alternative example. The acronym of these cells stands for "human mes-

encephalic neurons from fetal material from the university of Lund." This cell system allows measurement of functional endpoints and demonstrates the usefulness of image-based high-content screening in human cells. The cell system can be used for disease modeling (Lotharius et al., 2005) and also for examining toxicity to developed neurons (Schildknecht et al., 2009; Pörtl et al., 2012). In addition, it allows developmental neurotoxicity to be addressed (Stiegler et al., 2011). LUHMES are conditionally-immortalized neuronal precursor cells with dopaminergic neuronal features. Differentiation is triggered by a tetracycline-mediated inactivation of the v-myc transgene in the cells, and it results in uniformly post-mitotic neurons forming a neurite network (Fig. 2) (Scholz et al., 2011). Neurite outgrowth can be measured in an imaging-based procedure. Live imaging allows the simultaneous evaluation of cell viability and neurite outgrowth within one culture dish. Some compounds can slow the extension of neuronal processes at lower concentrations than those causing cell death. Extension-promoting compounds have been identified as well. To evaluate the specificity of the assay, the actions of unspecific cytotoxicants have been tested. Thus described test system may be useful for high-throughput screens to identify neurotoxic agents and for closer characterization concerning mode of action (MoA), compound interactions, or the reversibility of their effects (Stiegler et al., 2011).

As another example for toxicological use, the LUHMES cell system was used to examine the effects of methamphetamine or 1-methyl-4-phenylpyridinium (MPP⁺) and the parental compound 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). As expected, cells were sensitive to MPP⁺, while no reaction occurred with MPTP. The high homogeneity and purity of the cultures allowed the detection of metabolic changes during the degeneration, which is not possible in mixed primary human cell populations (Hansson et al., 2000). Cellular ATP in LUHMES cells dropped in two phases, cellular glutathione (GSH) decreased continuously, paralleled by an increase in lipid peroxidation. These events were accompanied by a time-dependent degeneration of neurites. Blocking the dopamine transporter completely abrogated MPP⁺ toxicity (Schildknecht et al., 2009). By applying different inhibitors, the underlying mechanisms and pathways have been identified. ATP-depletion, as the initial mitochondrial effect of MPP⁺, requires further downstream processes to result in neuronal death. These processes may form self-enhancing signaling loops that aggravate an initial energetic impairment and eventually determine cell fate (Pörtl et al., 2012; Schildknecht et al., 2011).

Further investigations are facilitated by genetic engineering of the cells. These manipulations can change the susceptibility to chemicals, test mechanisms of toxicity, or provide additional information in the form of reporter assays. Such approaches are a counterpart to the use of transgenic animals, and LUHMES have been used for many such approaches.

The intention of modeling the real world situation as closely as possible makes it necessary to place cells in contact with other cell types. Neuronal networks and tissues naturally consist of different cell types. Therefore, to predict reactions in humans these conditions must be reproduced experimentally, and

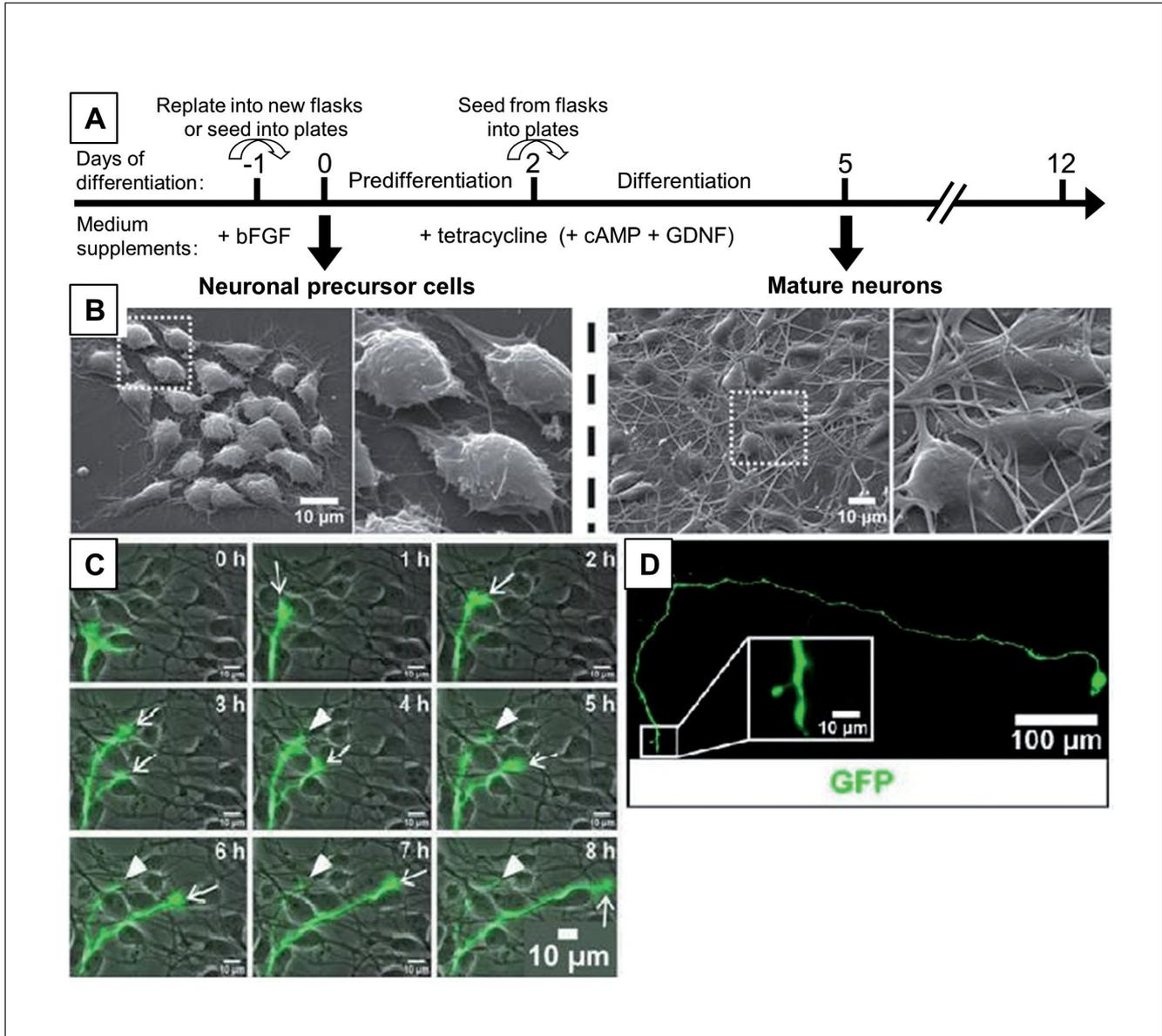


Fig. 2: LUHMES differentiation and neurite outgrowth during differentiation

Proliferating LUHMES cells can be amplified and easily converted into post-mitotic neurons. (A) Schematic representation of the 2-step differentiation procedure, initiated by withdrawal of the cytokine basic fibroblast growth factor (bFGF) and addition of tetracycline.

(B) Representative scanning electron microscopy (SEM) images of undifferentiated (day 0) and differentiated (day 5) LUHMES cells.

Marked squares indicate areas that are shown at higher magnification to the right. (C) Neurite outgrowth was documented by time-lapse microscopy from day 3 to day 5. A representative sequence of merged phase contrast and GFP channel images, recorded for 8 h on day 4 is shown. Expanding neurites with active growth cones are indicated by arrows, retracting neurites are marked by arrowheads.

(D) GFP-over-expressing "mature" day 5 LUHMES cell with a typical neurite length of 800 μm and an ending without growth cone (no further neurite growth). Insert shows neurite ending at higher magnification. Adapted from (Scholz et al., 2011).

possibly further refined by *in silico* modeling. One first step are co-cultures, as described already by Schildknecht et al. (2012). As astrocytes serve numerous functions, such as nutrient supply of neurons, regulation of cerebral blood flow, orchestration of neuronal growth and differentiation, maintenance of extracellular glutamate levels, and ion and liquid balance (Kettenmann and Ransom, 2005), they seem to be the right partners to support neurons in their biological function.

The future of this cell culture system lies in the combination of cell culture techniques and innovative analytical approaches. Metabolomics and transcriptomics are especially promising tools for providing a richer set of data. Combination with other cells and 2D or 3D structuring of the cultures appear very attractive. The lack of functional N-methyl-D-aspartate receptors is, however, a disadvantage compared to several primary neuron cultures (Volbracht et al., 2006). Thus, it is

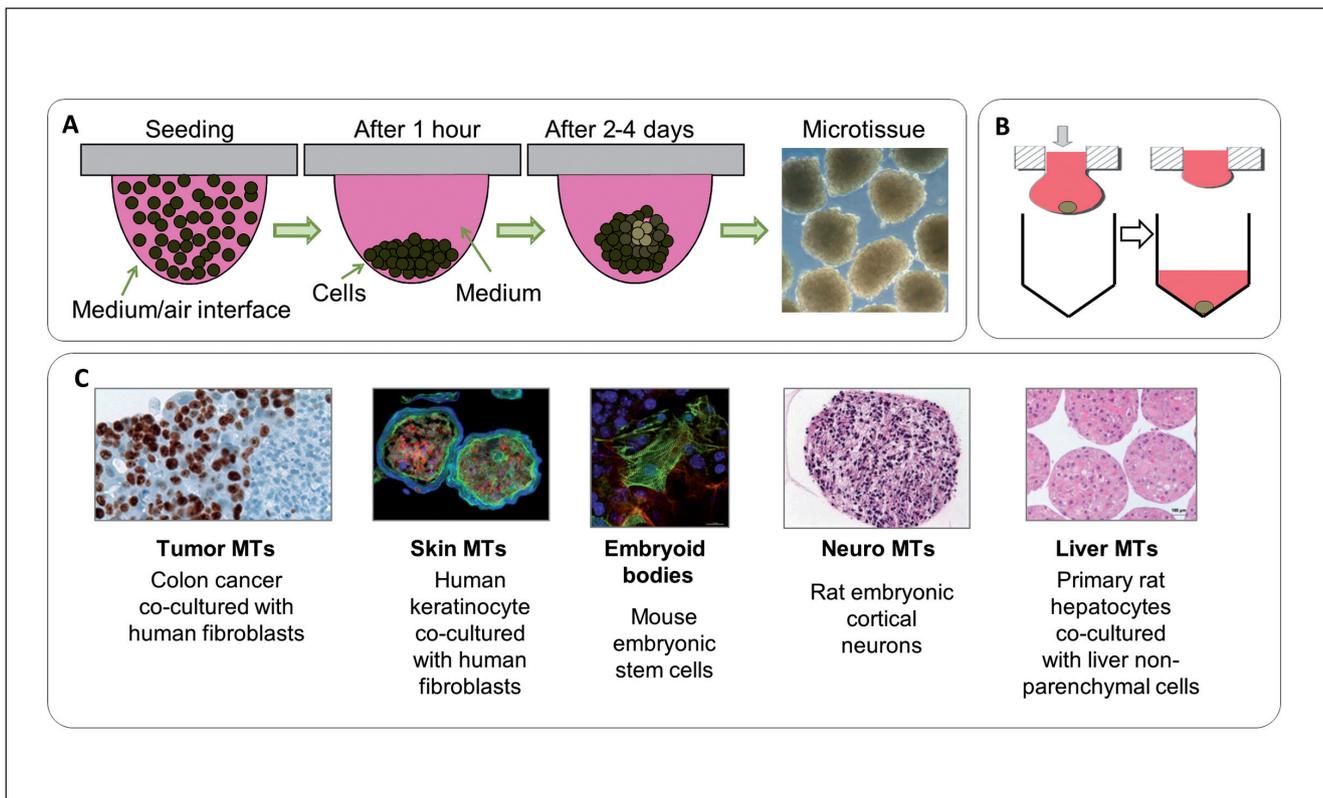


Fig. 3: Microtissue (MT) formation in hanging drops: overview of different microtissue models produced with the hanging drop technology

(A) The cells accumulate at the bottom of the drop, form cell-cell contacts and extracellular matrix and assemble to spherical microtissues. Formation of microtissues requires between 2-7 days depending on the model. (B) Automated production of microtissues in the GravityPLUS™ platform. After the formation process, microtissues are transferred into the microtissue assay plate (GravityTRAP™) for further cultivation and assay. (C) Overview of different microtissue models produced with the hanging drop technology.

important to know both the advantages and the limitations of LUHMES cells.

4 Automation-compatible organotypic microtissues for drug and substance testing

Physiological tissue-like culture systems seem to improve the relevance of *in vitro* testing of substances tremendously. Although the advantages of organotypic 3D cell culture models to increase the performance of *in vitro* compound assessment have been known for years (Pampaloni et al., 2007; Justice et al., 2009), complex production and elevated readout processes impeded the industrial implementation. Microtissue models can be derived either from cell lines, primary cells, or stem cells. They are spherical in shape and devoid of artificial biomaterials such as hydrogels or scaffold materials. The cells produce their own, cell type-specific extracellular matrix environment, mimicking native tissues *in vitro*. Often, they display some morphological tissue-like features, but they do not necessarily reflect the overall histology as seen *in vivo*. Nevertheless, biochemical functions of cells in 3D microtissues can be closer to the *in vivo* situation than to the one observed in 2D *in*

vitro cultures. As an example of a 3D liver microtissue model, the InSphero system was chosen, as this technology platform allows relatively high throughput and adaptation to many different cell types and requirements.

A novel automation-compatible 96-well platform based on hanging drops to produce organotypic microtissues has been developed (Kelm and Fussenegger, 2004), and spherical microtissues can be produced in a scalable production process. This is achieved by a well design where a microfluidic channel connects an inlet funnel at the top and an outlet funnel at the bottom of the plate, allowing for hanging drop formation without turning the plate upside down. At the well outlet a hanging drop is formed by a combination of capillary and surface-tension forces. In this the cells assemble into a tissue-like arrangement (Fig. 3). Microtissues can be initiated with low cell numbers resulting in excellent size uniformity with variations below 10% in diameter (Drewitz et al., 2011). The size is defined by the initial cell number. In contrast to tumorigenic and ES cells, primary cells are contact inhibited and do not proliferate (Drewitz et al., 2011).

Complementing the manufacturing platform for reproducible production of single and multi-cell type microtissues, a spheroid-specific culture plate allows more convenient long-term

cultures, medium exchanges, optical analysis, and biochemical assays. The lack of artificial biomaterials ensures compatibility with most biochemical assays used to assess cell viability and toxicity. The whole system can be combined with robotic liquid-handling devices equipped with a 96-multichannel pipette head, with similar volumetric precision as in standard multi-well plates (Drewitz et al., 2011). Future developments may include the combination of periportal and perivenous types of hepatocytes and closer modeling of the liver lobule structure, while at present the focus of the platform is on robustness, throughput and improved overall biochemical function.

An example application is the use of liver microtissues from primary rat hepatocytes or cryopreserved human hepatocytes, both in co-culture with non-parenchymal cells (Godoy, in preparation; Messner et al., submitted) for toxicity testing. In standard sandwich culture systems, liver-specific functionality decreases rapidly with prolonged culture time. The 3D environment already improves the functionality, whereas heterotypic cell populations incorporating NPCs further increase liver-specific functions as shown for urea secretion over time (Fig. 4). Actually, levels of secreted albumin are close to values detected for human native liver (Meng, 2010; Uygun et al., 2010). Incorporation of liver-derived macrophages enables the detection of indirect liver toxic effects such as inflammation-mediated toxicity (Messner et al., submitted).

Microtissues are a versatile 3D cell culture concept to create a wide variety of different tissue models, which are designed for high-throughput data generation and easy implementation in current drug development processes to foster drug de-risking.

The key advantages of microtissue models are:

- Long term maintenance of tissue structure and functionality
- Multi-cell type models
- No scaffold requirements (no impact of batch to batch variations from biomaterials and hydrogels)
- Production of endogenous cell type-specific extracellular matrix
- Direct cell-cell and cell-ECM interactions
- Standard 96-well format enables automation and high-throughput compatibility
- Same tissue format for efficacy and safety assessment

5 Use of genetically engineered 3D models for identification of toxicity pathways

3D organotypic *in vitro* human epithelial models are important advances over traditional monolayer cell culture models. For toxicology applications, these models provide a number of important features (Kandárová et al., 2009; Kaluzhny et al., 2011).

Specific advantages of 3D organotypic human epithelial models include:

- Normal (non-immortal) human cells
- Organotypic structure
- Barrier function
- Real-life exposure options

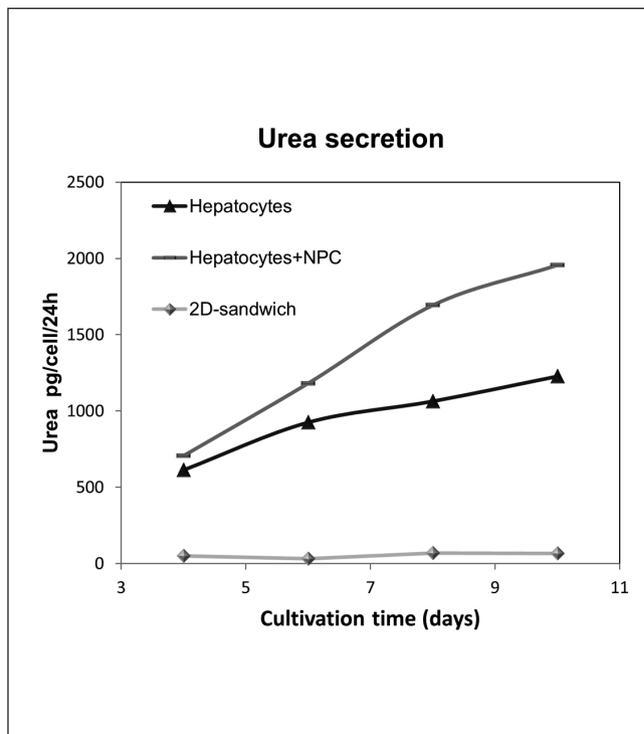


Fig. 4: Urea secretion of primary rat hepatocytes

Rat hepatocytes were grown in 2D sandwich cultures, or as 3D microtissues. The microtissues either contained non-parenchymal liver cells (NPC) as additional cell populations or they were formed from hepatocyte monocultures. The cells were kept in culture for 10 days, and the urea production was measured every second day, starting from day 4. The data were normalized to the number of cells in culture, and are given as production per day.

- Xenobiotic metabolizing capabilities
 - Co-culture models for epithelial-stromal interactions
- 3D human skin (e.g., EpiDerm™) and corneal tissue (e.g., EpiOcular™) models have become increasingly important as replacements for traditional animal-based toxicology testing in the area of cosmetics, personal care products, household cleaning products, and in the chemical and pharmaceutical industries. Two EpiDerm™-based test methods for assessing skin irritation and dermal corrosion potential are now formally validated as alternative methods in the European Union (EU). Several other EpiDerm™ and EpiOcular™ test methods continue through the validation process as defined by regulators in the EU (ECVAM) and the United States (ICCVAM).

Normal human 3D (NHu-3D) epithelial models with the added feature of engineered toxicological reporter functions are currently being developed (Hayden et al., 2011). These models represent a further advance that will allow development of mechanistic toxicity screening assays. Initial experiments have produced promising results.

Early passage normal human epidermal keratinocytes, dermal fibroblasts, tracheobronchial epithelial cells, and fibroblasts were transduced with lentiviral vectors containing NFκB reporters linked to either GFP or luciferase. Stably transduced

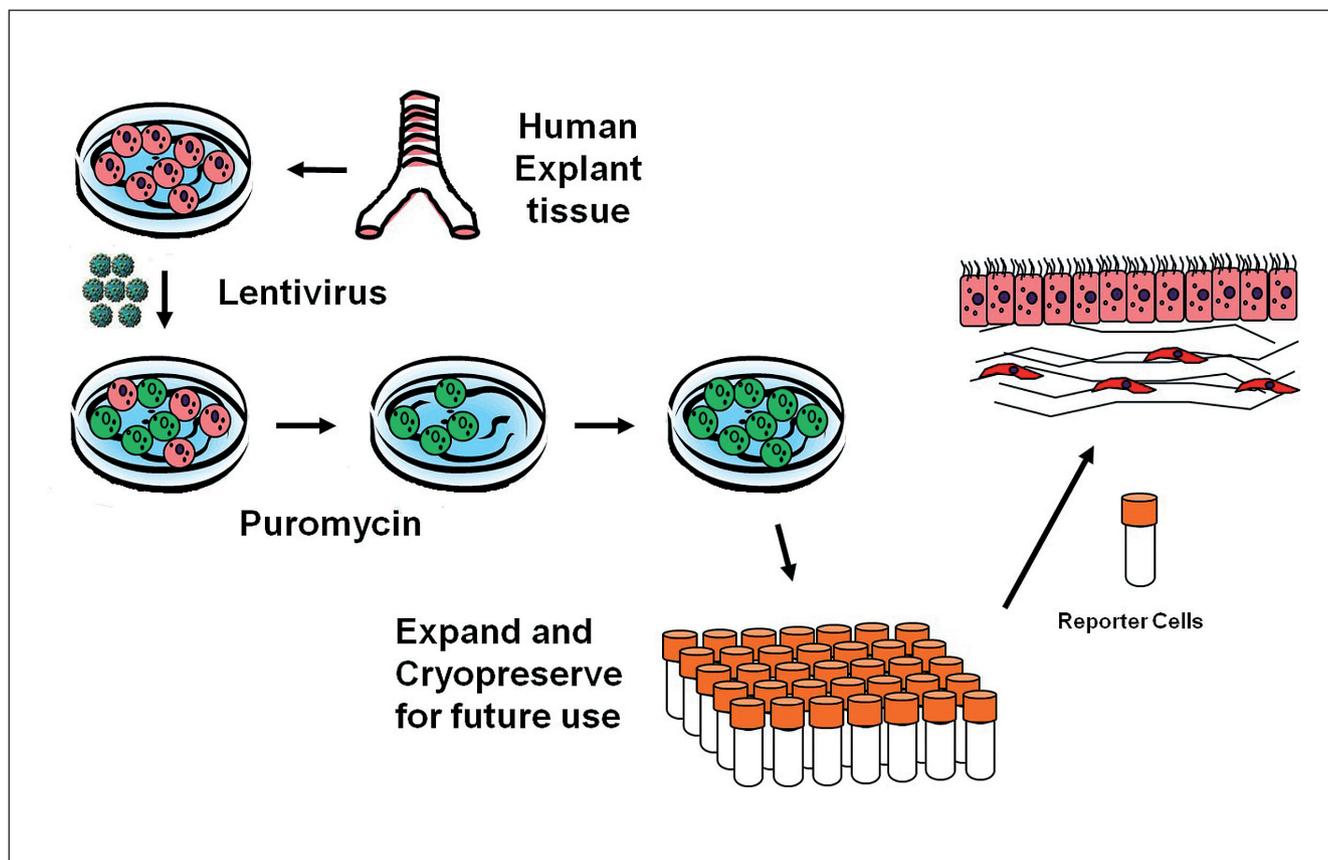


Fig. 5: 3D organotypic reporter models from normal human cells

Schematic representation of the process for creation of organotypic reporter models from normal human cells. Normal cells are isolated from human explants (e.g., skin or lung) and kept in culture (shown in red). They are transduced with lentiviruses containing reporter constructs and an antibiotic resistance. The successfully transduced cells (shown in green) are selected with puromycin, expanded and cryopreserved for long-term storage. After recovery, cryopreserved cells are utilized for production of organotypic reporter models.

cells were selected by puromycin resistance, expanded over several passages and cryopreserved to produce large pools of reporter-expressing cells (Fig. 5). Reporter-expressing cells were then utilized to produce NHu-3D skin and airway epithelial models.

Important considerations for the development of organotypic reporter models are:

- Stable integration of reporter into cells
- Attainment of normal organotypic development
- Attainment of adequate reporter activity
- Generation of adequate number of cells to support commercial production
- Simple, robust assays for HTP screening

Organotypic structure and barrier properties of models produced from reporter-expressing cells were found to be similar to models produced from untransduced cells, as determined by histological and barrier assessment. NFκB reporters linked to either GFP or luciferase were found to be activated approximately 5-fold above background by treatment of the organotypic models with tumor necrosis factor (TNF)α. GFP was detected in formalin fixed paraffin sections by epifluorescence microscopy, and luciferase activity in tissue extracts was quan-

tified with a microplate luminometer (Fig. 6). Production of models containing other reporters of toxicological significance (e.g., for DNA damage, oxidative stress, heavy metal stress, ER stress, etc.) by the same process will provide a suite of human epithelial reporter models that can be utilized to provide mechanistic toxicity screening assays.

6 Prediction of skin irritancy using mechanism-based integration of *in vitro* and *in silico* methods

In the field of skin irritation, numerous test models have been developed over the years, some of which have been validated and accepted by OECD as stand-alone alternative methods to the Draize rabbit skin test (OECD, 2010). *In silico* methods also are available, including rule-based expert systems and QSAR prediction of binary classification of skin irritancy. For example, ToxTree implemented a decision tree for skin irritation rules (Hulzebos et al., 2005) that requires users to provide physicochemical properties as exclusion rules prior to applying inclusion rules based on structural alerts.

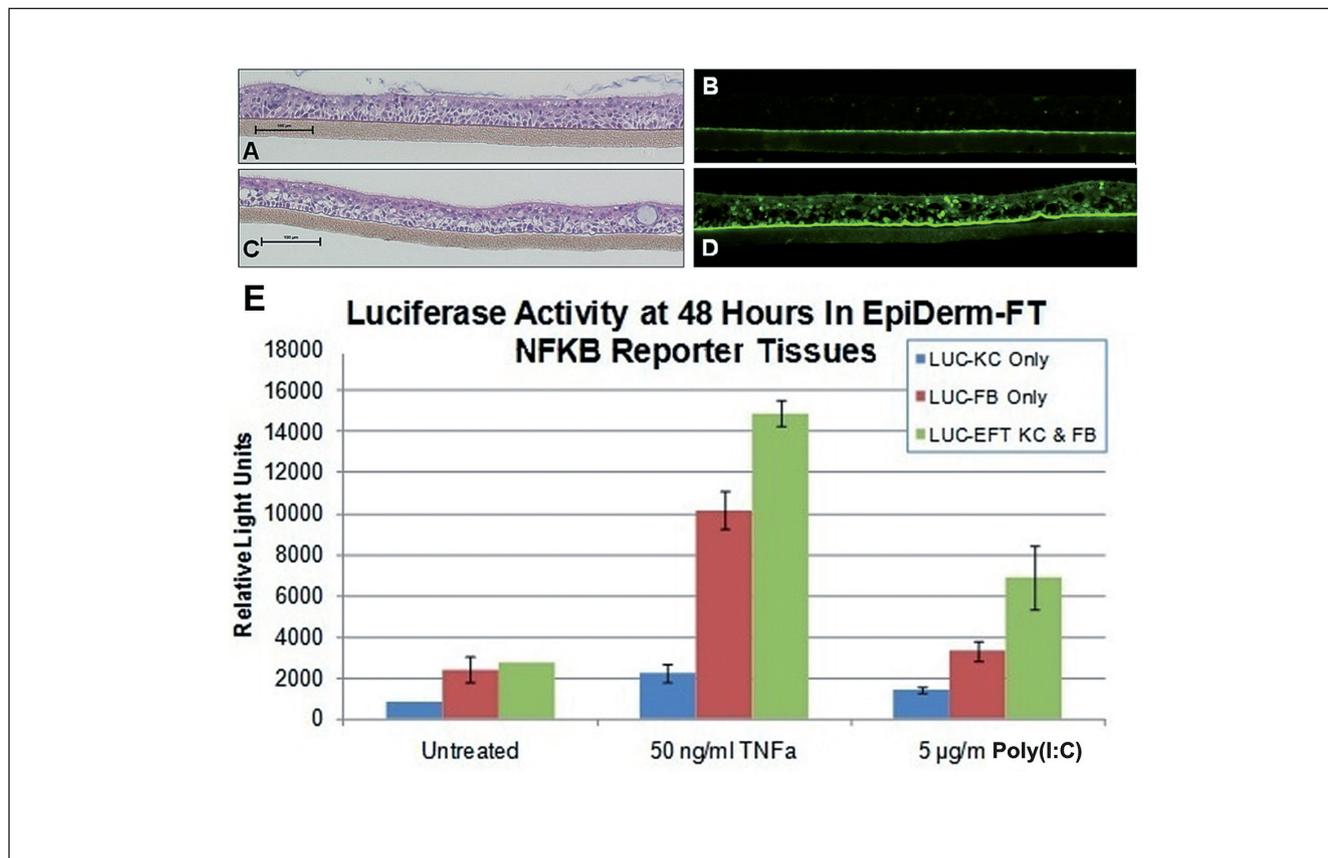


Fig. 6: Activation of EpiDerm-FT™ NFκB reporter by TNF α or Poly(I:C)

The human EpiDerm-FT™ skin model is shown as an example for reporter tissues. EpiDerm tissue was generated from cells containing reporter constructs for NFκB (green fluorescent protein expression under the control of NF-κB for microscopic assays; luciferase expression for quantitative biochemical assays). Control tissue (A) displayed typical cornified squamous epithelial morphology, and only the basal membrane showed a fluorescence signal (autofluorescence, B). After stimulation with 50 ng/ml TNF α for 48 h, tissue morphology appeared normal (C). Visualization of the reporter signal by fluorescence microscopy showed the activation of the NFκB reporter in TNF α treated tissue (D). (E) For quantitative assessment of EpiDerm-FT™ NFκB reporter activity, tissue was generated with fibroblasts (FB) containing reporter constructs, or with keratinocytes (KC) containing reporter constructs or with both cell types containing reporter constructs. The tissues were then stimulated either with TNF α or with the toll-like receptor agonist polyinosinic acid (Poly(I:C)). After 48 h, the tissue was harvested, homogenized and used for determination of luciferase (luc) activity.

Here we present an automated workflow algorithm developed and implemented to predict a molecule's potential for skin irritancy (including severity grading). *In vivo* Draize legacy data (OECD, 2002) have been used as learning and training data for the model.

MoA QSAR modeling approach

A theoretical framework to link initial interactions of a chemical with its target site to toxicologically relevant endpoints on the level of the organism or the population are the adverse outcome pathways (AOP), defined initially in the field of ecotoxicology (Ankley et al., 2010). The concept has been taken up by several authorities dealing more broadly with environmental and chemical risk assessment, and it is further elaborated and applied to case studies, for instance by the OECD. An AOP is identified by a causal linkage between a molecular initiating event (MIE) and *in vivo* endpoint of regulatory value for risk

assessment. Levels in between may include cellular changes and responses of tissues or organ systems. Before AOPs were defined, the related concept of mode-of-action (MoA) was in use for a long time. It also describes the chain of events, explaining how a chemical leads to functional effects on a higher level of complexity (cellular or organism level). In many cases, the term MoA is used to describe the initial parts of an AOP at a level of high biochemical resolution.

A MoA QSAR modeling approach may be useful to reflect mechanisms when constructing training sets and selecting model descriptors. The MoA QSAR approach includes the preparation of mechanistic training sets by identifying MoA categories that are linked to the phenotypic effects. MoA categories are defined based on chemical groups that initiate the molecular interactions in putative toxicity pathways and result in biological events that lead to phenotypic effects. The chemical groups that initiate this chain of events and therefore link

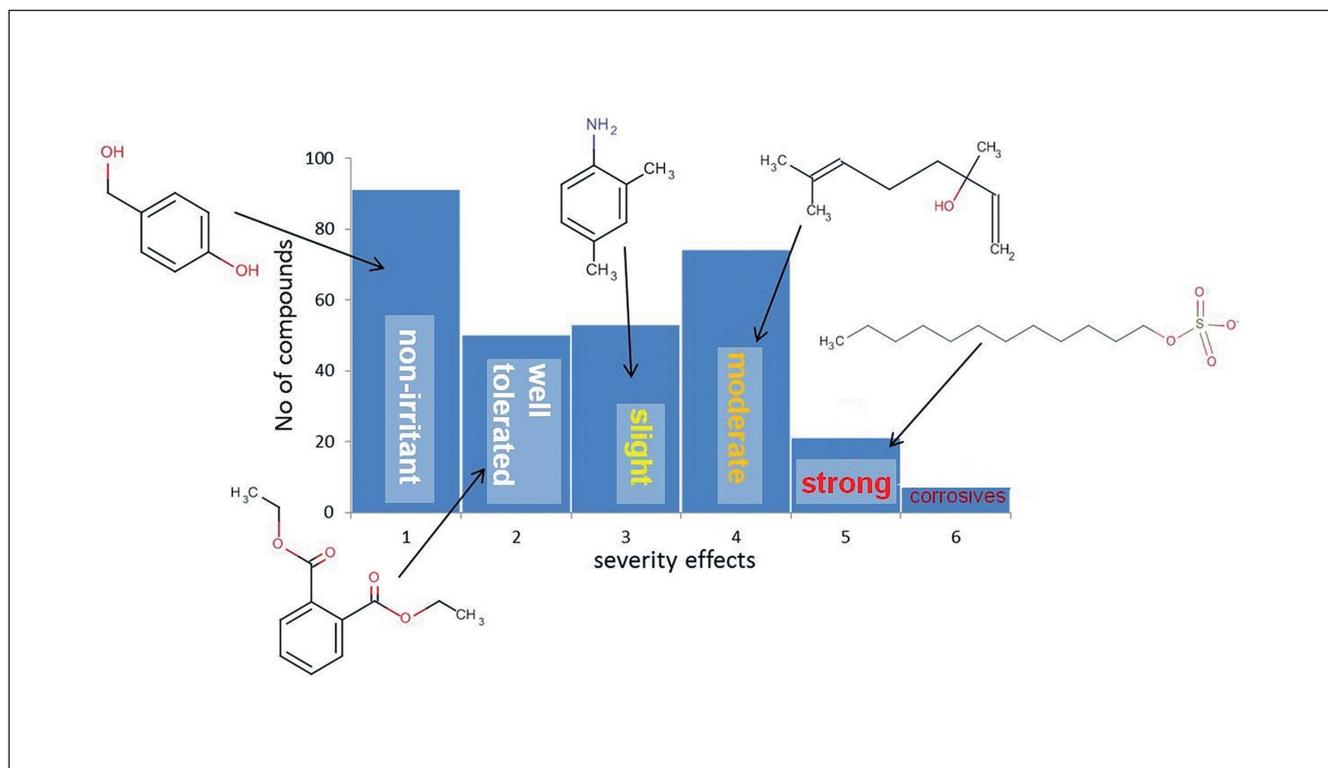


Fig. 7: Distribution of skin irritation severity effects in the training set

Distribution of the severity of skin irritation in the training set according to L'Oréal criteria: non-irritant (87 structures), well tolerated (42), slight (45), moderate (68), and strong (8).

biological event pathways to phenotypic effects are defined as chemical MoA categories. These are identified from individual, highly resolved MoA pathways by aggregating several pathways into one category. The purpose of this “pooling” step was to build large groups of chemicals containing both positives and negatives in the training set. These chemical groups include surfactants, acids and bases (alkali and bleaches), organic solvents, and reactive groups. From a L'Oréal internal source and the ECETOC report (ECETOC, 1995), a collection of 269 chemical structures were selected as the global dataset. This global set was then partitioned into individual MoA groups: alcohols (98 structures), alkenes (68), amines (50), carboxylic acids/esters (93), reactive groups (135), and surfactants (84) (Fig. 7).

Here the surfactant group is selected as an example and treated in detail. Surfactants can act in a number of ways, including the triggering of inflammation and/or cell death (De Jongh et al., 2006), altering signaling pathways (Torma and Berne, 2009), or enhancing the loss of barrier function by disrupting lipid bilayer structures (Welss et al., 2004). While the former are widely studied by *in vitro* assays and gene/mRNA expressions, the membrane damaging can be investigated by applying colloid chemistry and *in physico* assays (Yang et al., 2001). Surfactant molecules can disrupt lipid bilayers in two ways, and the purpose of these *in physico* assays is to develop descriptors to be used in our MoA modeling process.

Modeling procedure

Molecular descriptors were used to relate aspects of chemical structures to biological responses. Structural features were represented as fingerprints from a fragment library of generic features, as well as structural alerts for skin irritation. Physicochemical properties included calculated LogP (lipophilicity), water solubility, polarizability, polar surface area, and dipole moment. Further parameters that were used include molecular shape descriptors (e.g., diameter, moment of inertia) and surfactant-specific parameters, such as HLB (hydrophilic-lipophilic balance) and molecular packing factors (Israelachvili, 1994). Based on the structural features and properties, in the first part of the approach, a binary classification of non-irritants vs. irritants was performed for each MoA group as well as the global dataset. This was done using a combination of partial least squares (PLS) and logistic regression. The probabilities of being an irritant were then combined from the different models into one overall outcome using a quantitative weight of evidence (WoE) approach. Each model is assigned a weight to optimize the prediction of the non-irritants during the training stage.

$$p_{final} = \frac{\sum w_i p_i}{\sum w_i}$$

where w_i is the weight of each model and p_i is the probability of being an irritant for a given chemical. The sensitivity was 95% and the specificity was 84%.

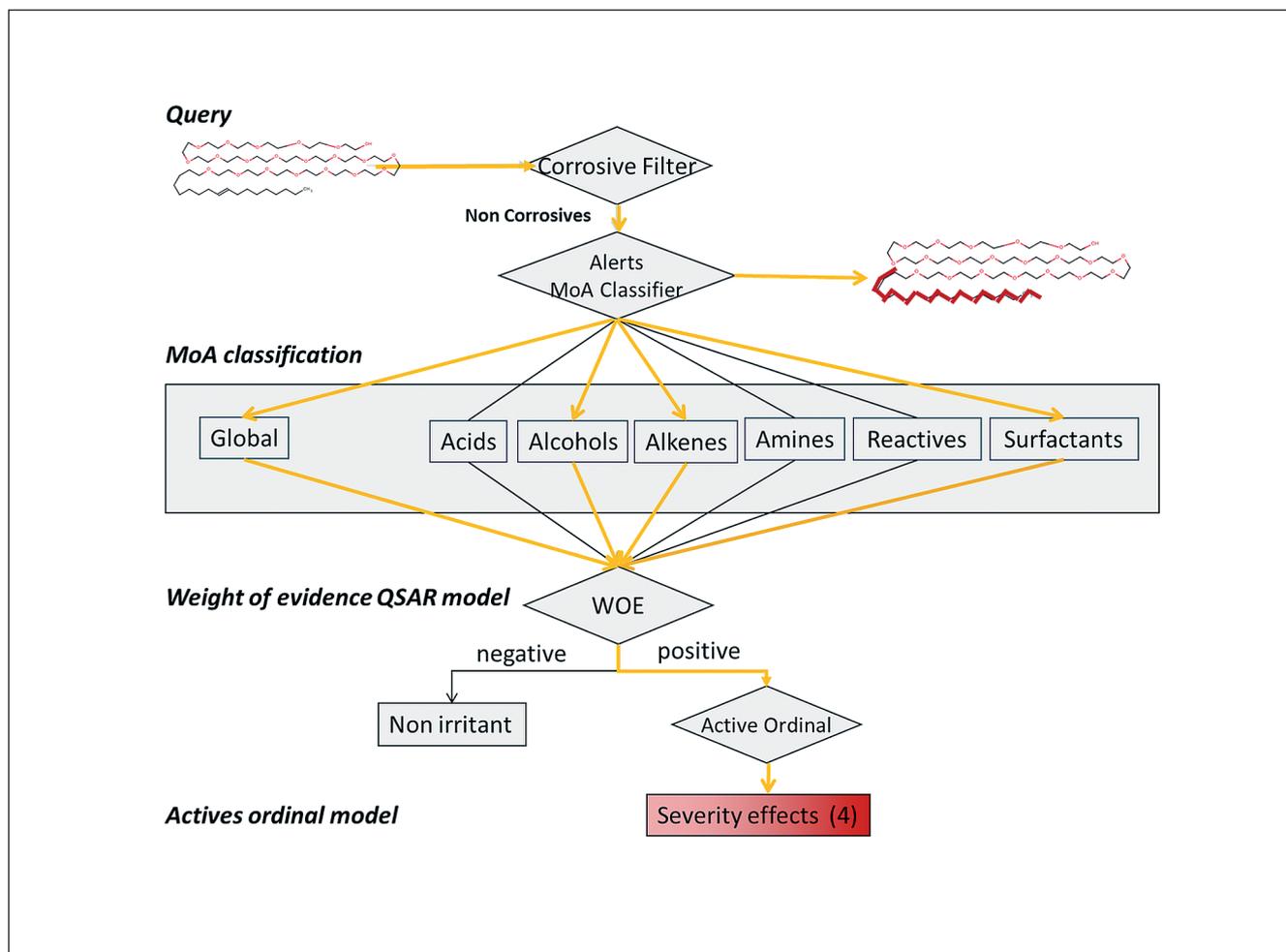


Fig. 8: Decision tree for the MoA QSAR

The applicability domain of the model is evaluated based on both structural features and physicochemical properties used in the models. MoA: mode of action; WOE: weight of evidence.

Next, irritant chemicals are further ranked for severity effects using ordinal partial least squares regression. The applicability domain of the model was evaluated based on both structural features and physicochemical properties used in the models (Fig. 8).

As part of L'Oréal's integrated testing strategy, a workflow system (Fig. 9) was designed for easy and transparent access to the decision tree of structural alerts and MoA QSAR models for *in vivo* skin irritation.

7 Human variability: Solutions through the application of *in silico* pattern recognition and knowledge discovery methods to human data

Research undertaken with human subjects and samples suffers from large inter-individual variations, and therefore encourages the view that the science is potentially of poorer quality since predictions emanating from such human research will be less reliable and repeatable. Therefore, the issue of variation

dictates much of the ongoing preference for inbred animals in controlled laboratory environments as tools to understand human health and disease, both from “quality science” and experimental control perspectives. This approach largely neglects the problem of species differences (Olson et al., 2000; Hartung, 2008), when the main purpose of toxicological research is the extrapolation to humans. In modern toxicological test systems, this issue is not fully resolved when conclusions for humans are drawn from zebra fish, drosophila, or cultures of non-human cells. The species extrapolations are made particularly difficult, as “man” is not one genetically defined entity. Large variations, relevant for toxicological and disease outcome, are being mapped by the ENCODE project, for instance. A different approach is taken by the example technology presented here: improved pattern recognition in heterogeneous human data sets, in order to more sharply define human responses on a broad population level.

Variation issues associated with the data produced from human studies are a significant challenge in developing human models suitable as replacement alternatives for animals in biomedical,

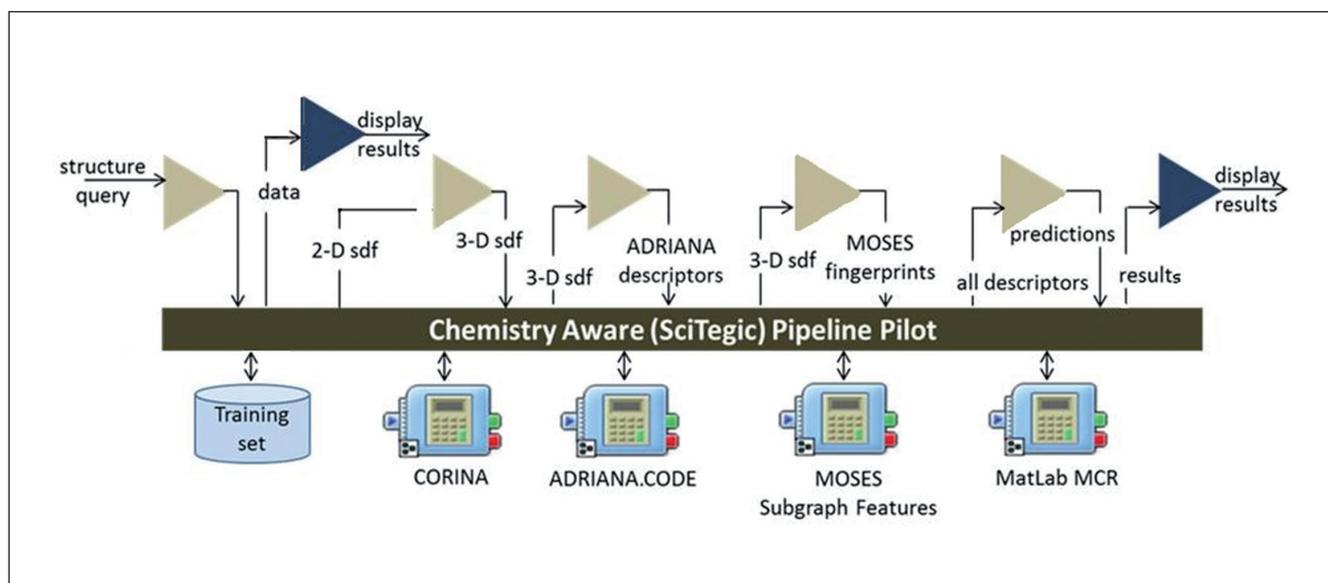


Fig. 9: Workflow implementation of “Skin Irritation Predictor”

Essential components of the system configured and aligned through a chemistry-aware pipelining technology.

toxicological, and other health research and investigations. Additional impetus for considering alternatives to animals comes through an increasing focus on the unreliability of translation from some animal models to human disease, and thereafter the development of vaccines and therapeutics (Pound et al., 2004; Hackam and Redelmeier, 2006; Khanna and Burrows, 2011).

Human variability

An advantage from modern medical and health systems is the generation of enormous volumes of data. Large data sets lend themselves to modern computational approaches, such as machine learning and sophisticated multi-dimensional statistical techniques that together comprise the bioinformatics field of pattern recognition. Machine learning methods, for example recursive partitioning (decision trees) and support vector machines (SVM), are particularly attractive, since algorithms like these can be trained to identify patterns in highly variable data associated with a response (e.g., test for a viral infection, drug side-effects), allowing more accurate predictions from complex data in the subsequent testing phase. These techniques, borrowed from computer science and statistics and combined with access to vast human data sets, provide another avenue through which to approach biological variation and provide an alternative to mice with experimenter-defined genetics and living environment. In short, to overcome the diversity of human subjects that can confound interpretation, sophisticated *in silico* pattern recognition methods can identify data profiles for sub-populations represented by the predictor variables analyzed and based on a biological or toxicological response of interest.

Machine learning examples: Decision trees and support vector machines (SVM)

Decision trees have been applied to clinical decision-making involving diverse and complex patient data (Crowley et al.,

2002), so the application of machine learning has precedent in the medical knowledge domain. A decision tree is a data classification method generated by asking serial questions on the features associated with data items. This can be simple “yes or no” questions contained within a node, with each node comprising a distinct yes or no outcome. From the top node (the root) a hierarchical path based on yes/no answers continues until a node without further outcomes (a “leaf”) is reached; a data item is thereafter assigned to this class (i.e., the terminal classification outcome) (Kingsford and Salzberg, 2008).

Also popular with *in silico* pattern recognition studies are SVM. SVM are very powerful for data classification and novelty detection, and they can be used for regression modeling. SVM have the advantage of a simple input space for data entry, with analysis occurring in a high-dimension feature space (Φ) defined by a kernel function. The basis of modeling is the kernel, the simplest example being the linear kernel, which represents the images of two data points (x, x') in the high-dimension feature space (1):

$$(1) k(x, x') = \langle \Phi(x), \Phi(x') \rangle \text{ (Karatzoglou et al., 2006)}$$

Depending on prior knowledge of the data (e.g., distribution, structure) and application, other kernel functions are available (Karatzoglou et al., 2006; Smola and Scholkopf, 2004).

The application of decision trees in tandem with SVMs is being developed to overcome the difficulty of human variability as a barrier to replacement alternatives for animal studies in biomedical research. This approach provides a pattern recognition tool for large multi-parameter (up to thirty predictor variables) data collections containing thousands of individual cases (i.e., patient results). Decision trees have the advantage of multiple decision boundaries that, when combined with SVMs, provide a powerful means to identify meaningful patterns in complex human health data, either prospectively or retrospectively, which can guide further investigations into a disease mechanism or toxic drug response.

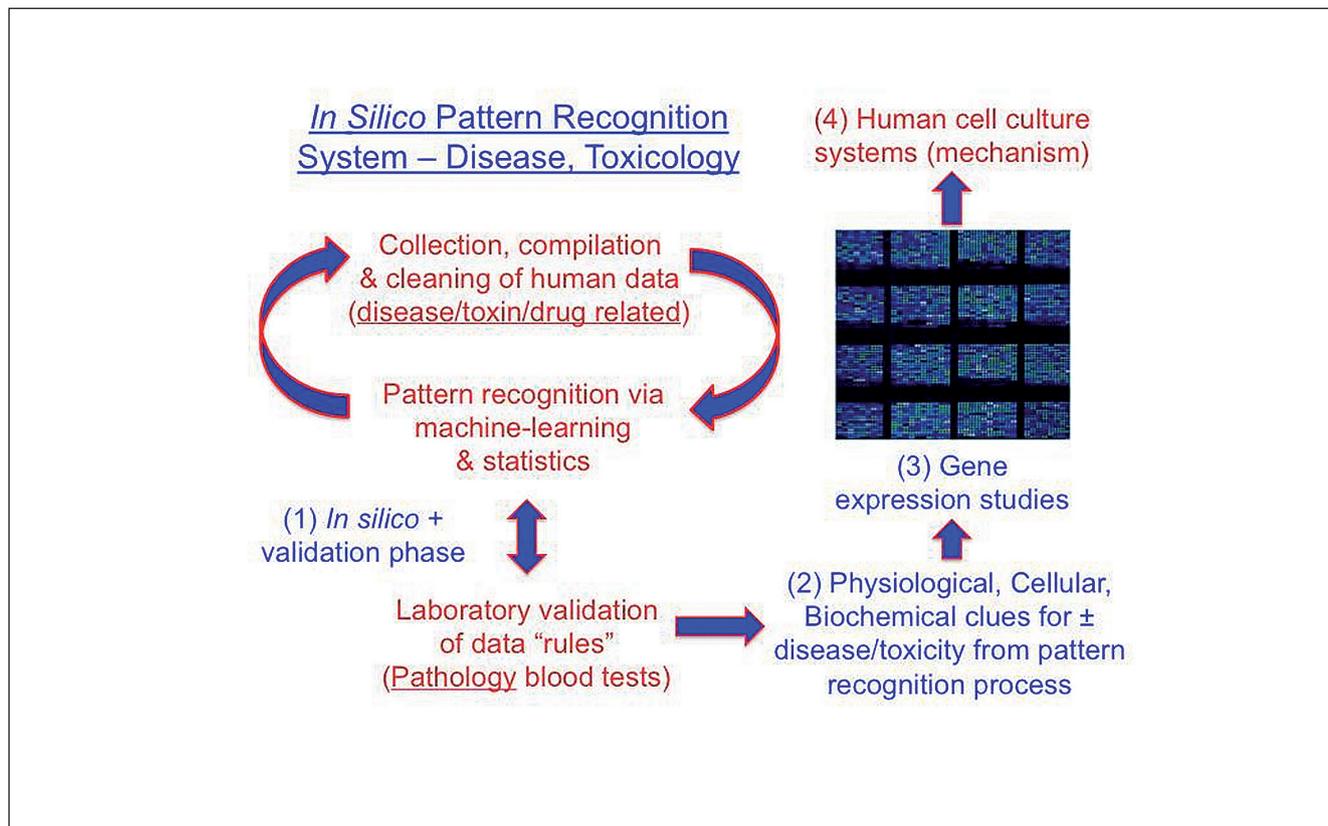


Fig. 10: Extraction of patterns from clinical/epidemiologic data despite large human variability

A strategy to overcome natural human variation based on *in silico* pattern recognition studies, linked to biological validation investigations as a replacement alternative for inbred animals in biomedical research and toxicology assessments. Further validation through genetic investigations also assists the identification of mechanisms that can be further explored by human cell 2D and 3D culture systems, such as those described elsewhere in this review. The individual phases of this strategy include (1) Data preparation and *in silico* pattern recognition modeling of complex bulk human data by data mining and machine learning, in tandem with biological validation of data patterns associated with a defined response (e.g., infection, drug toxicity); (2) Identification of data patterns associated with important physiological, cellular, or biochemical processes associated with disease or toxicity response, which will (3) guide the choice of gene expression chip/panel for deeper bioinformatics analysis and comparison; (4) Investigation of response mechanisms through human cell culture investigations, guided by findings from phase 3. Figure adapted from Lidbury et al. (2012).

A pattern recognition system, using only human data, is central to a new system that employs interaction of the wet laboratory for biological validation with computational pattern models and linkage to modern genetic analytical techniques and databases to further explore the basis of physiological clues detected in the *in silico* pattern recognition phase of investigations on aggregated pathology data (Fig. 10).

Natural variation in human genetics and environment need not be a barrier to fundamental biomedical research, toxicology studies, or any other aspect of human health research. With the massive accumulation of human data, whether through pathology testing, social and psychological analyses, or epidemiology, data analytic techniques borrowed from computer science and statistics make it possible to expose relevant and unique patterns within highly variable human populations. With such variation successfully harnessed and linked to biological validation strategies and sophisticated human genetic knowledge, viable alternatives to animal studies will emerge, particularly

when tethered to modern knowledge of human cell culture systems for mechanistic studies.

8 Conclusion

Each of the technical approaches and model systems presented here has been developed as a stand-alone method. Most have been developed for a specific purpose and to solve defined problems. Dozens, if not hundreds, of such technologies are already available, and only a few have been picked to exemplify the progress in the field. We have put forward the hypothesis here that added value may be generated by a combination of such approaches. The approach taken here may, at first glance, look different from or in competition with other new strategies. For instance, the ToxCast program, or different approaches that follow the “Tox21” vision take a different starting point. In their extreme form, they rid themselves of the old patchwork



of different toxicological models, be they *in vivo* or *in vitro* (Hartung and BcBride, 2011; NRC, 2007; Leist et al., 2008a), and put forward a new homogeneous framework, based, for instance, on PoT and systems biology modeling. It is not yet clear, which role assays play that use endpoints that are toxicologically apparently simple but (systems-) biologically highly complex, e.g., cell death, neurite degeneration, or albumin secretion.

Here we take an alternative approach to define an overall scaffold of what information would contribute to an animal-free risk assessment. This scaffold is used to recruit a largely heterogeneous group of assays, providing information at different levels of complexity, with different throughput rates, and possibly with different information value. Combined in a scheme, these assays can fill knowledge gaps and improve the overall risk assessment of chemicals for which little is known.

The framework suggested here is also suited to the incorporation of individual tests and *in silico* methods developed for Tox21, or even to incorporation of testing strategies at a higher level of integration, as shown by the Altamira example of skin irritancy modeling. Thus, this approach may represent a practical solution for high production volume risk assessment in the intermediate future, while many tests are still under development and no complete test platform on the basis of PoT testing is available. The future will then bring higher throughput assays, better systems biology modeling, better integration of data from omics technologies, and better cell sources. For instance, we envisage that testing in non-transformed cell models, of murine (Sipes et al., 2011a) or preferentially of human origin, will require a further development of stem cell technology, to provide reliable cell sources (Leist et al., 2008c; Zimmer et al., 2011a,b, 2012; Weng et al., 2012; Balmer et al., 2012).

References

- Ankley, G. T., Bennett, R. S., Erickson, R. J., et al. (2010). Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29, 730-741.
- Balmer, N. V., Weng, M. K., Zimmer, B., et al. (2012). Epigenetic changes and disturbed neural development in a human embryonic stem cell-based model relating to the fetal valproate syndrome. *Hum Mol Genet* 21, 4104-4114.
- Blaauboer, B. J. (2010). Biokinetic modeling and in vitro-in vivo extrapolations. *J Toxicol Environ Health B Crit Rev* 13, 242-252.
- Blaauboer, B. J., Boekelheide, K., Clewell, H., et al. (2012). The use of biomarkers of toxicity for integrating in vitro hazard estimates into risk assessment for humans. *ALTEX* 29, 411-425.
- Bolt, H. M., Roos, P. H., and Thier, R. (2003). The cytochrome P-450 isoenzyme CYP2E1 in the biological processing of industrial chemicals: consequences for occupational and environmental medicine. *Int Arch Occup Environ Health* 76, 174-185.
- Coecke, S., Pelkonen, O., Leite, S. B., et al. (2012). Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches. *Toxicol In Vitro*. doi: 10.1016/j.tiv.2012.06.012
- Crowley, S., Tognarini, D., Desmond, P., et al. (2002). Introduction of lamivudine for the treatment of chronic hepatitis B: expected clinical and economic outcomes based on 4-year clinical trial data. *J Gastroenterol Hepatol* 17, 153-164.
- De Jongh, C. M., Verberk, M. M., Withagen, C. E., et al. (2006). Stratum corneum cytokines and skin irritation response to sodium lauryl sulfate. *Contact Dermatitis* 54, 325-333.
- Drewitz, M., Helbling, M., Fried, N., et al. (2011). Towards automated production and drug sensitivity testing using scaffold-free spherical tumor microtissues. *Biotechnol J* 6, 1488-1496.
- ECETOC (1995). TR 066: Skin irritation and corrosion: Reference Chemicals Data Bank. <http://www.ecetoc.org/technical-reports>
- Falsig, J., Latta, M., and Leist, M. (2004). Defined inflammatory states in astrocyte cultures: correlation with susceptibility towards CD95-driven apoptosis. *J Neurochem* 88, 181-193.
- Godoy, P. (in preparation). Liver in vitro systems in pharmacology, toxicology and basic research. "The hepatosphere model".
- Hackam, D. G. and Redelmeier, D. A. (2006). Translation of research evidence from animals to humans. *JAMA* 296, 1731-1732.
- Hansson, O., Castilho, R. F., Kaminski Schierle, G. S., et al. (2000). Additive effects of caspase inhibitor and lazaroide on the survival of transplanted rat and human embryonic dopamine neurons. *Exp Neurol* 164, 102-111.
- Hartung, T. (2008). Thoughts on limitations of animal models. *Parkinsonism Relat Disord* 14, Suppl 2, S81-83.
- Hartung, T. and Hoffmann, S. (2009). Food for thought ... on in silico methods in toxicology. *ALTEX* 26, 155-166.
- Hartung, T. and McBride, M. (2011). Food for thought ... on mapping the human toxome. *ALTEX* 28, 83-93.
- Hayden, P. J., Jackson, G. R., Stolper, G., et al. (2011). Organotypic in vitro human epithelial models with engineered toxicologic reporter functions. *The Toxicologist* 120, Suppl 2., Abstract 2839.
- Hengstler, J. G., Marchan, R., and Leist, M. (2012). Highlight report: towards the replacement of in vivo repeated dose systemic toxicity testing. *Arch Toxicol* 86, 13-15.
- Hulzebos, E., Walker, J. D., Gerner, I., and Schlegel, K. (2005). Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. *QSAR Comb Sci* 24, 332-342.
- IPCS – International Programme on Chemical Safety (2004). IPCS glossary of international exposure assessment terminology. In: *IPCS Risk Assessment Terminology*. Geneva: World Health Organization (Harmonization Project Document No. 1).
- Israelachvili, J. (1994). The science and applications of emulsions – an overview. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 91, 1-8.
- Judson, R. S., Kavlock, R. J., Setzer, R. W., et al. (2011). Estimating toxicity-related biological pathway altering doses for



- high-throughput chemical risk assessment. *Chem Res Toxicol* 24, 451-462.
- Justice, B. A., Badr, N. A., and Felder, R. A. (2009). 3D cell culture opens new dimensions in cell-based assays. *Drug Discov Today* 14, 102-107.
- Kadereit, S., Zimmer, B., van Thriel, C., et al. (2012). Compound selection for in vitro modeling of developmental neurotoxicity. *Front Biosci* 17, 2442-2460.
- Kaluzhny, Y., Kandarova, H., Hayden, P., et al. (2011). Development of the EpiOcular™ eye irritation test for hazard identification and labelling of eye irritating chemicals in response to the requirements of the EU cosmetics directive and REACH legislation. *Altern Lab Anim* 39, 339-364.
- Kandárová, H., Hayden, P., Klausner, M., et al. (2009). An in vitro skin irritation test (SIT) using the EpiDerm reconstructed human epidermal (RHE) model. *J Vis Exp* 29, e1366.
- Karatzoglou, A., Meyer, D., and Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software* 15, 1-28.
- Kavlock, R., Chandler, K., Houck, K., et al. (2012). Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem Res Toxicol* 25, 1287-1302.
- Kelm, J. M. and Fussenegger, M. (2004). Microscale tissue engineering using gravity-enforced cell assembly. *Trends Biotechnol* 22, 195-202.
- Kettenmann, H. and Ransom, B. R. (eds.) (2005). *Neuroglia* (2nd edition). New York, USA: Oxford University Press.
- Khanna, R. and Burrows, S. R. (2011). Human immunology: a case for the ascent of non-furry immunology. *Immunol Cell Biol* 89, 330-331.
- Kingsford, C. and Salzberg, S. L. (2008). What are decision trees? *Nat Biotechnol* 26, 1011-1013.
- Kirkland, D., Reeve, L., Gatehouse, D., and Vanparys, P. (2011). A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. *Mutat Res* 721, 27-73.
- Kuegler, P. B., Baumann, B. A., Zimmer, B., et al. (2012). GFAP-independent inflammatory competence and trophic functions of astrocytes generated from murine embryonic stem cells. *Glia* 60, 218-228.
- Latta, M., Künstle, G., Leist, M., and Wendel, A. (2000). Metabolic depletion of ATP by fructose inversely controls CD95- and tumor necrosis factor receptor 1-mediated hepatic apoptosis. *J Exp Med* 191, 1975-1985.
- Leist, M., Hartung, T., and Nicotera, P. (2008a). The dawning of a new age of toxicology. *ALTEX* 25, 103-114.
- Leist, M., Kadereit, S., and Schildknecht, S. (2008b). Food for thought ... on the real success of 3R approaches. *ALTEX* 25, 17-32.
- Leist, M., Bremer, S., Brundin, P., et al. (2008c). The biological and ethical basis of the use of human embryonic stem cells for in vitro test systems or cell therapy. *ALTEX* 25, 163-190.
- Leist, M., Hasiwa, N., Daneshian, M., and Hartung, T. (2012). Validation and quality control of replacement alternatives current status and future challenges. *Toxicol Res* 1, 8-22.
- Lidbury, B. A. and Richardson, A. M. (2012). A pattern recognition bioinformatics alternative system to rodent models in fundamental research. *ALTEX Proceedings* 1, 515-520.
- Lotharius, J., Falsig, J., van Beek, J., et al. (2005). Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. *J Neurosci* 25, 6329-6342.
- Meng, Q. (2010). Three-dimensional culture of hepatocytes for prediction of drug-induced hepatotoxicity. *Expert Opin Drug Metab Toxicol* 6, 733-746.
- Messner, S., Agarkova, L., Moritz, W., and Kelm, J. M. (submitted). Multi-cell type human liver microtissues for hepatotoxicity testing.
- Mirlohi, S., Dietrich, A. M., and Duncan, S. E. (2011). Age-associated variation in sensory perception of iron in drinking water and the potential for overexposure in the human population. *Environ Sci Technol* 45, 6575-6583.
- NRC (2007). *Toxicity testing in the 21st century: a vision and a strategy*. Washington, DC: National Academy Press.
- OECD (2002). Test No. 404: Acute Dermal Irritation/Corrosion. http://www.oecd-ilibrary.org/environment/test-no-404-acute-dermal-irritation-corrosion_9789264070622-en
- OECD (2010). Test No. 439: In Vitro Skin Irritation Reconstructed Human Epidermis Test Method. <http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/OECD/OECD-TG439.pdf>.
- Olson, H., Betton, G., Robinson, D., et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56-67.
- Padilla, S., Corum, D., Padnos, B., et al. (2012). Zebrafish developmental screening of the ToxCast Phase I chemical library. *Reprod Toxicol* 33, 174-187.
- Pampaloni, F., Reynaud, E. G., and Stelzer, E. H. (2007). The third dimension bridges the gap between cell culture and live tissue. *Nat Rev Mol Cell Biol* 8, 839-845.
- Pohl, H. R. and Scinicariello, F. (2011). The impact of CYP2E1 genetic variability on risk assessment of VOC mixtures. *Regul Toxicol Pharmacol* 59, 364-374.
- Pörtl, D., Schildknecht, S., Karreman, C., and Leist, M. (2012). Uncoupling of ATP-depletion and cell death in human dopaminergic neurons. *Neurotoxicology* 33, 769-779.
- Pound, P., Ebrahim, S., Sandercock, P., et al. (2004). Where is the evidence that animal research benefits humans? *BMJ* 328, 514-517.
- Raunio, H. (2011). In silico toxicology non-testing methods. *Front Pharmacol* 2, 33.
- Rotroff, D. M., Dix, D. J., Houck, K. A., et al. (2012). Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. *Environ Health Perspect*, doi:10.1289/ehp.1205065.
- Schildknecht, S., Pörtl, D., Nagel, D. M., et al. (2009). Requirement of a dopaminergic neuronal phenotype for toxicity of low concentrations of 1-methyl-4-phenylpyridinium to human cells. *Toxicol Appl Pharmacol* 241, 23-35.
- Schildknecht, S., Pape, R., Müller, N., et al. (2011). Neuroprotection by minocycline caused by direct and specific scavenging of peroxynitrite. *J Biol Chem* 286, 4991-5002.
- Schildknecht, S., Kirner, S., Henn, A., et al. (2012). Characteri-



- zation of mouse cell line IMA 2.1 as a potential model system to study astrocyte functions. *ALTEX* 29, 251-265.
- Scholz, D., Pörtl, D., Genewsky, A., et al. (2011). Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *J Neurochem* 119, 957-971.
- Sipes, N. S., Martin, M. T., Reif, D. M., et al. (2011a). Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. *Toxicol Sci* 124, 109-127.
- Sipes, N. S., Padilla, S., and Knudsen, T. B. (2011b). Zebrafish: as an integrative model for twenty-first century toxicity testing. *Birth Defects Res C Embryo Today* 93, 256-267.
- Smola, A. J. and Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing* 14, 199-222.
- Stiegler, N. V., Krug, A. K., Matt, F., and Leist, M. (2011). Assessment of chemical-induced impairment of human neurite outgrowth by multiparametric live cell imaging in high-density cultures. *Toxicol Sci* 121, 73-87.
- Torma, H. and Berne, B. (2009). Sodium lauryl sulphate alters the mRNA expression of lipid-metabolizing enzymes and PPAR signalling in normal human skin in vivo. *Exp Dermatol* 18, 1010-1015.
- Uygun, B. E., Soto-Gutierrez, A., Yagi, H., et al. (2010). Organ reengineering through development of a transplantable recellularized liver graft using decellularized liver matrix. *Nat Med* 16, 814-820.
- Valerio, L. G., Jr. (2009). In silico toxicology for the pharmaceutical sciences. *Toxicol Appl Pharmacol* 241, 356-370.
- Valerio, L. G., Jr. (2011). In silico toxicology models and databases as FDA Critical Path Initiative toolkits. *Hum Genomics* 5, 200-207.
- van Thriel, C., Westerink, R. H., Beste, C., et al. (2012). Translating neurobehavioural endpoints of developmental neurotoxicity tests into in vitro assays and readouts. *Neurotoxicology* 33, 911-924.
- Volbracht, C., van Beek, J., Zhu, C., et al. (2006). Neuroprotective properties of memantine in different in vitro and in vivo models of excitotoxicity. *Eur J Neurosci* 23, 2611-2622.
- Welss, T., Basketter, D. A., and Schroder, K. R. (2004). In vitro skin irritation: facts and future. State of the art review of mechanisms and models. *Toxicol In Vitro* 18, 231-243.
- Weng, M. K., Zimmer, B., Pörtl, D., et al. (2012). Extensive transcriptional regulation of chromatin modifiers during human neurodevelopment. *PLoS One* 7, e36708.
- Yang, C., Ansong, C., Bockrath, L. J., et al. (2001). Mechanistic study of model monolayer membranes and their interactions with surfactants: correlation to effects on CHO cell cultures. In N. O. Yasuhiro Iwasawa and K. Hironobu (eds.), *Studies in Surface Science and Catalysis*. Elsevier.
- Zimmer, B., Kuegler, P. B., Baudis, B., et al. (2011a). Coordinated waves of gene expression during neuronal differentiation of embryonic stem cells as basis for novel approaches to developmental neurotoxicity testing. *Cell Death Differ* 18, 383-395.
- Zimmer, B., Schildknecht, S., Kuegler, P. B., et al. (2011b). Sensitivity of dopaminergic neuron differentiation from stem cells to chronic low-dose methylmercury exposure. *Toxicol Sci* 121, 357-367.
- Zimmer, B., Lee, G., Balmer, N. V., et al. (2012). Evaluation of developmental toxicants and signaling pathways in a functional test based on the migration of human neural crest cells. *Environ Health Perspect* 120, 1116-1122.

Acknowledgements

This work was supported by the Doerenkamp-Zbinden Foundation and by grants from the EU FP7 program (ESNATS), EU FP7 FETopen program "The body on a chip," No. 296257. CY and JFR acknowledge funding from L'Oréal and the COSMOS consortium (EU FP7 and Cosmetics Europe No. 266835 in SEURAT). PJH and coworkers were funded by the US National Institute of General Medical Sciences (NIGMS), and BAL is supported by a fellowship and funding from The Medical Advances without Animals Trust (MAWA) (<http://www.mawa-trust.org.au/>).

Correspondence to

Marcel Leist, PhD
University of Konstanz
Box M657
78457 Konstanz
Germany
e-mail: marcel.leist@uni-konstanz.de