Mutat Res Gen Tox En 850–851 (2020) 503135

Contents lists available at ScienceDirect

Mutat Res Gen Tox En

journal homepage: www.elsevier.com/locate/gentox

Use of *in vitro* 3D tissue models in genotoxicity testing: Strategic fit, validation status and way forward. Report of the working group from the 7th International Workshop on Genotoxicity Testing (IWGT)

Stefan Pfuhler[a,*], Jan van Benthem[b], Rodger Curren[c], Shareen H. Doak[d], Maria Dusinska[e], Makoto Hayashi[f], Robert H. Heflich[g], Darren Kidd[h], David Kirkland[i], Yang Luan[j], Gladys Ouedraogo[k], Kerstin Reisinger[l], Toshio Sofuni[m], Frédérique van Acker[n], Ying Yang[o], Raffaella Corvi[p]

[a] *Procter and Gamble, Mason Business Centre, Mason, OH, USA*
[b] *National Institute for Public Health and the Environment, Centre for Health Protection, Bilthoven, the Netherlands*
[c] *Institute for In Vitro Sciences, Inc., Gaithersburg, MD, USA*
[d] *Swansea University Medical School, Singleton Park, Swansea, SA2 8PP, Wales, UK*
[e] *Health Effects Laboratory, Department of Environmental Chemistry, NILU-Norwegian Institute for Air Research, Kjeller, Norway*
[f] *makoto international consulting, Ebina, Japan*
[g] *U.S. Food and Drug Administration/National Center for Toxicological Research, Jefferson, AR, USA*
[h] *Covance Laboratories Ltd, Otley Road, Harrogate, HG3 1PY, UK*
[i] *Kirkland Consulting, PO Box 79, Tadcaster, LS24 0AS, UK*
[j] *School of Public Health, Hongqiao International Institute of Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, PR China*
[k] *L'Oréal R&I, Aulnay-sous-bois, France*
[l] *Henkel AG & Co KGaA, Duesseldorf, Germany*
[m] *Formerly National Institute of Health Sciences, Tokyo, Japan*
[n] *Triskelion B.V., Zeist, the Netherlands*
[o] *Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, PR China*
[p] *European Commission, Joint Research Centre (JRC), Ispra, Italy*

ARTICLE INFO

Keywords:
In vitro genotoxicity testing
DNA damage
Comet assay
Micronucleus test
Reconstructed human skin
Liver spheroids
3D airway models

ABSTRACT

Use of three-dimensional (3D) tissue equivalents in toxicology has been increasing over the last decade as novel preclinical test systems and as alternatives to animal testing. In the area of genetic toxicology, progress has been made with establishing robust protocols for skin, airway (lung) and liver tissue equivalents. In light of these advancements, a "Use of 3D Tissues in Genotoxicity Testing" working group (WG) met at the 7th IWGT meeting in Tokyo in November 2017 to discuss progress with these models and how they may fit into a genotoxicity testing strategy. The workshop demonstrated that skin models have reached an advanced state of validation following over 10 years of development, while liver and airway model-based genotoxicity assays show promise but are at an early stage of development. Further effort in liver and airway model-based assays is needed to address the lack of coverage of the three main endpoints of genotoxicity (mutagenicity, clastogenicity and aneugenicity), and information on metabolic competence. The IWGT WG believes that the 3D skin comet and micronucleus assays are now sufficiently validated to undergo an independent peer review of the validation study, followed by development of individual OECD Test Guidelines.

1. Introduction

Use of three-dimensional (3D) tissue equivalents in toxicology has been increasing over the last decade as novel preclinical test systems and as alternatives to animal testing [1–3]. In the area of genetic toxicology, progress with establishing robust protocols has been made for skin, airway and liver tissue equivalents. In light of these advancements, a "Use of 3D Tissues in Genotoxicity Testing" working group (WG) was formed that met at the 7th IWGT meeting in Tokyo, November 2017.

* Corresponding author.
E-mail address: pfuhler.s@pg.com (S. Pfuhler).

https://doi.org/10.1016/j.mrgentox.2020.503135
Received 9 January 2020; Accepted 9 January 2020
Available online 15 January 2020

Previous expert panels discussed the potential use and status of these assays and helped set the stage for the 2017 W G meeting. More specifically, at the 2009 IWGT meeting the WG “In Vitro Genotoxicity Test Approaches With Better Predictivity” discussed the best developed and most frequently used models based on human 3D reconstructed skin (RS). This WG agreed that RS-based genotoxicity models, once validated, will be useful to follow up on positive results from standard *in vitro* assays for dermally applied compounds [4]. It became clear, however, that more work was needed to ensure a robust model and the testing of more coded chemicals was recommended, as well as further evaluation of the metabolic capacity of the RS models. A broader discussion about the use of *in vitro* tissue equivalents, going beyond skin models, was held at the "New Technologies” Workshop in 2012 in Washington D.C., USA, hosted by the Health and Environmental Sciences Institute’s Genetic Toxicology Technical Committee (HESI-GTTC) [5]. The strengths and weaknesses of 3D skin, airway and liver models were discussed at the Workshop. The Workshop report acknowledged that the ‘*in vivo*-like’ behavior of 3D tissue constructs was an important advantage of these models and recommended them as superior in this regard to the standard 2D static cell culture systems, which they concluded were artificial and far removed from the *in vivo* state. Conversely, it was noted that 3D tissue-based assays are more technically difficult to perform, more expensive, and have a lower throughput than assays conducted with 2D cell cultures. It was also noted that, at that time, assays with 3D cultured systems were used only in a small number of laboratories.

At the IWGT in Tokyo, a diverse WG, comprised of representatives from regulatory institutions, academia and industry from Asia, North America and Europe, was chartered with the task of reviewing recent progress with the development, optimization and validation of *in vitro* tissue models for genotoxicity testing. The WG consisted of subject matter experts, some of whom were bringing experimental data to the WG to enable efficient and evidence-based discussions towards the specific WG goals, which were:

- Review the available genotoxicity data generated in liver, airway and skin 3D tissue models
- Discuss the validation status of these assays and their fit in a genotoxicity testing strategy
- Develop recommendations for further development of these assays and capture consensus statements

2. Experimental data presented/discussed

2.1. ‘3D’ liver models in genotoxicity testing

Liver spheroids can be readily constructed from hepatocytes and one such model has been developed using HepG2 hepatocellular carcinoma cells, based upon growth in a hanging-drop format [6]. When HepG2 cells are cultured in this 3D spheroid format, they exhibit a substantial increase in liver-specific functionality, expressing significantly higher CYP1A1/2 activities and production of albumin and urea than the same cells grown in 2D monolayer format [6]. A protocol for utilizing 3D HepG2 liver spheroids with the cytokinesis blocked micronucleus (CBMN) assay has been established. To evaluate their suitability for genotoxicity testing in the CBMN assay, 3D HepG2 spheroids were exposed to the pro-carcinogens, benzo[*a*]pyrene (BaP) and 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP). The performance of the standard CBMN assay in HepG2 cell monolayers (2D) was then compared to the 3D HepG2 spheroid assay [6]. Following exposure to BaP, the first significant increase ($p < 0.05$) in micronucleus (MN) frequency occurred at 3 μM BaP in both 2D monolayers and the 3D HepG2 hanging-drop spheroids; however, the MN frequency induced by this concentration of BaP in the 3D hanging-drop spheroids was 2-fold higher than in 2D monolayers. With respect to PhIP, the lowest observable effect level occurred at a lower concentration when the 3D liver spheroids were exposed to the compound, as compared to the standard 2D test system. The lowest concentration that resulted in a significant induction of micronuclei in the 3D hanging-drop spheroids was 5 μM PhIP, while in the 2D HepG2 culture system, it was 10 μM PhIP. Furthermore, the level of genotoxicity induced by this lowest observable effect level in the 3D spheroids was nearly 2-fold higher than in the standard 2D CBMN assay. Thus, both BaP and PhIP exhibited significantly higher MN frequencies at the same concentrations in the 3D models than in the standard 2D monolayer cultures of HepG2 cells. Additionally, PhIP was positive for genotoxicity at a lower concentration in the 3D model than in the 2D CBMN assay [6]. This higher level of chromosomal damage in the 3D CBMN assay with both BaP and PhIP is thought to be due to the higher metabolic activity exhibited in the HepG2 cells when cultured in a spheroid format, which more efficiently converts these compounds into their genotoxic metabolites.

2.2. ‘3D’ airway models in genotoxicity testing

Human 3D airway models (sometimes referred to as lung models) consist of fully differentiated and functional human respiratory epithelium, including cilia, mucus layer, etc., and allow relevant exposure to air as they are cultured at an air-liquid interface (ALI). These models are formed from primary cultures of human airway epithelial cells (typically from the large airway of donor lungs at autopsy) which are allowed to differentiate at the ALI over a period of approximately one month to form a mixture of basal, ciliated, goblet and possibly club cells, forming a layer of cells that closely resembles the lining of the human airway [7]. It is anticipated that these models may enable a more realistic (geno)toxicity assessment of inhaled compounds; and unlike the RS models, these tissue equivalents remain stable in culture for months, opening the possibility of using subchronic treatments similar to those used *in vivo*.

A protocol for the comet assay was established using two commercially available human reconstructed 3D airway models (MucilAir™ produced by Epithelix Sàrl, Switzerland and EpiAirway™ produced by MatTek Corporation, US) and one model developed in-house (LUMC, The Netherlands). Background levels of DNA damage in both of the commercially available and in the in-house untreated models were low, and concentration-related responses were observed following treatment with various well known genotoxins, such as methyl methane sulfonate (MMS), 4-nitroquinolone-*N*-oxide (4-NQO) and cyclophosphamide (CP) (Fig. 1). Positive responses for DNA damage using the comet assay have also been reported recently for 4-(methylnitrosamine)-1-(3-pyridyl)-1-butanone (NNK) by the US FDA/NCTR using a human ALI airway culture system developed in-house [8]. In addition to the comet assay, a protocol for the CBMN assay was applied. After several experiments to establish conditions for generating slides of good quality (with respect to cell density, nuclei and cytoplasm), it was concluded that the rate of cell division was too low to obtain a sufficient number of binucleated cells, making the detection of micronuclei less sensitive (Fig. 2).

The performance of the comet assay in a commercially available human 3D airway model, MucilAir™, was compared to the routinely used bronchial epithelial cell line BEAS-2B and the tumor cell line A549, which resemble alveolar pneumocyte Type II cells [9]. Upon receipt, the MucilAir™ models were maintained in culture (on 24-well Transwell™ culture supports) at the ALI using MucilAir™ culture medium. For air–liquid exposures of the A549 and BEAS-2B cells, cells were seeded onto track-etched polyethylene terephthalate (PET) membrane inserts. After 72 h, the inserts were ‘air-lifted’ (i.e., culture medium was removed from the apical side of the inserts) and the cells were cultured for 16–24 h before exposure with the apical surface exposed to air, while the basal surface was fed with medium through the membrane support. Cerium oxide nanoparticles (CeO_2; primary particle size 13.8 nm) were applied to the apical surface of the cultures via a dynamic airflow using the Vitrocell® system. The response of the different cell types upon exposure to an air stream (clean air) was

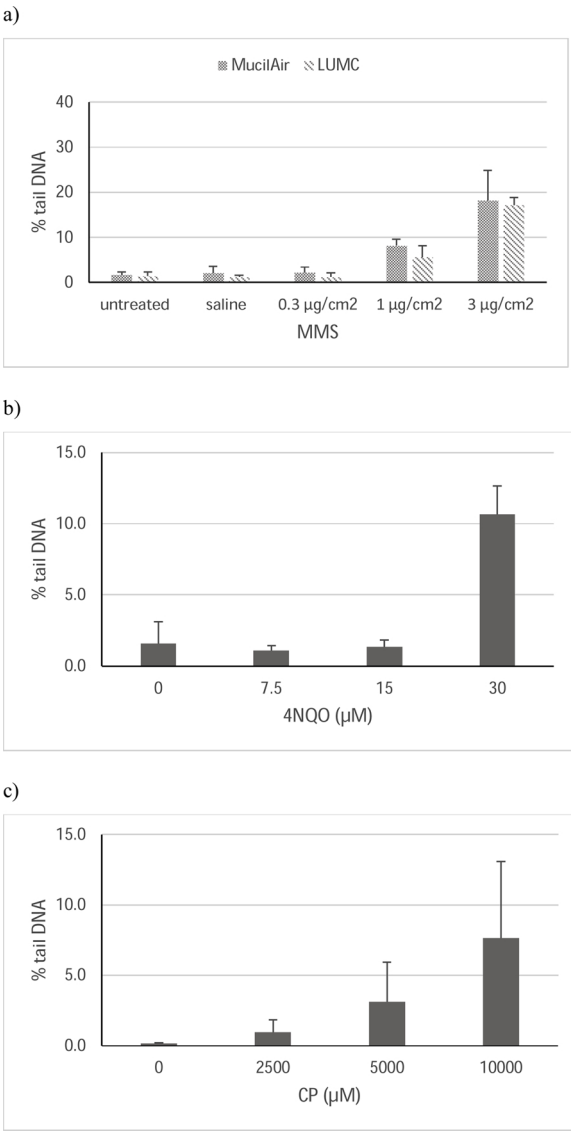

Fig. 1. a-c: Examples of concentration responses obtained with MMS, 4NQO and CP in 3D airway models.

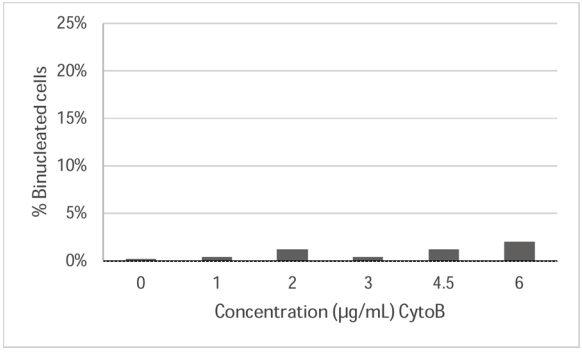

Fig. 2. Percentage binucleated cells after incubation of MucilAir™ with different concentrations of cytochalasin B.

determined by LDH leakage (Fig. 3A) and IL-8 release (Fig. 3B) and expressed relative to the incubator controls (not exposed to an air stream). Exposure to air alone did not increase the LDH and IL-8 responses in MucilAir™ compared to incubator controls, but in both BEAS-2B and A549 cells, it resulted in significant increases [9]. This is most

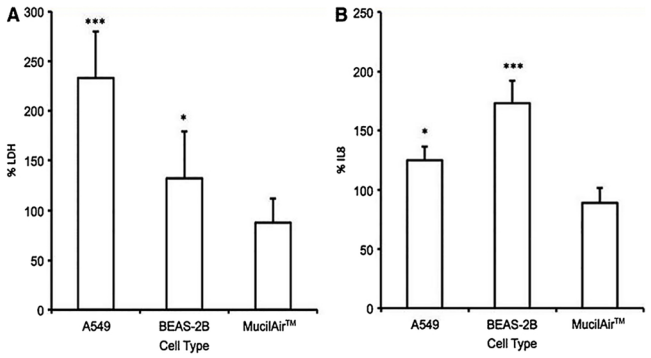

Fig. 3. LDH (A) and IL-8 (B) response of A549, BEAS-2B, and MucilAir™ to air exposure (without any CeO2). The biological LDH (A) and IL-8 (B) response (y axis) of the cells tested is expressed as %LDH and %IL-8 of air-exposed samples (n = 3) divided by the average expression of incubator control samples (n = 2). Indicated are the average and standard deviation of the resulting three ratios multiplied by 100 to obtain percentage values. */*** = statistically significant ($0.01 < p < 0.05$ / $p < 0.0001$). From Kooter et al. [9] Mary Ann Liebert, Inc., publishers, with permission.

likely due to the protective morphology of the ALI culture, having cilia, mucus layer, etc. similar to the lining of the human airway. Aerosolizing CeO_2 resulted in agglomeration or aggregation of particles. Such agglomerates or aggregates are likely to be the predominant form of particle that interact with all types of cells. Only 14 % of particles were below 100 nm and the mean particle size was 300 nm. The deposited concentrations of CeO_2 particles for the A549 and BEAS-2B cells were 0.04, 0.16, and 0.71 µg/cm^2. MucilAir™ models received 0.15, 0.67, and 3.0 µg/cm^2. Cellular responses to exposure via air to different concentrations of CeO_2 for 1 h varied according to cell type tested. Exposure of the MucilAir cultures did not result in significant effects in terms of inflammatory or cytotoxicity parameters and genotoxicity (comet assay), but it did affect the cell lines. Conversely, oxidative stress (increased Heme oxygenase 1 protein expression) was observed in the MucilAir™ cultures but not in the cell lines. This suggests that the human 3D airway models may predict a more realistic, *in-vivo*-like response; whereas, 2D cultures might overestimate a potentially toxic effect of nanoparticles.

2.3. '3D' skin models in genotoxicity testing

The current status of the validation efforts for the reconstructed skin micronucleus test (RSMN) and RS comet assay was presented. Experimental data were shown from an international validation effort that started in 2006, with support from Cosmetics Europe (CE) and the German Federal Ministry for Education and Research (BMBF). This validation project is part of a strategy of the CE Genotoxicity Task Force towards developing an *in-vitro*-only genotoxicity testing strategy for cosmetic ingredients [10,11]. This project was initiated as a result of the 7th Amendment to the EU Cosmetics Directive, which bans *in vivo* genotoxicity testing for cosmetics, effective since 2009 [12]. Addressing limitations of the skin assays discussed at the 5th IWGT in Basel [4], the CE project has added more coded chemicals to the validation dataset and also has investigated the metabolic competency of commercially available 3D human RS skin models, specifically EpiDerm™ [13] and Phenion® FT [14,15]. It was found that the metabolic competency of these RS models is similar to native human skin [13,14], thereby confirming that RS skin models have *in-vivo*-like metabolic properties, consistent with their use as '2nd tier' assays to follow-up on positive results from standard 2D testing battery assays. Advantageously, the use of comet and MN RS assays allows the investigation of all key modes of genotoxic activity mandatory for regulatory testing (albeit in the form of an indicator assay), which should help improve the sensitivity of the follow-up test.

2.3.1. The reconstructed skin comet assay (RS comet), validation data and examples of regulatory use

The state of RS comet assay protocol development [16] was discussed by the WG, with presentation of data from the first two phases of the validation exercise. In addition, the final protocol used for the validation study was presented, including criteria for assay validity and data evaluation. Phenion® Full-Thickness (FT) skin tissues were used for the validation, although the general suitability of another FT skin tissue (EpiDerm FT™) also has been shown [16]. For FT tissues consisting of keratinocytes and fibroblasts, the protocol foresees analyzing both cell types with the alkaline version of the comet assay to identify a wide spectrum of DNA damage. In the standard protocol, tissues are treated by repeated dosing at 48, 24 and 3 h before dissociation of the tissues; in a second experiment, this protocol was amended by the addition of the DNA repair inhibitor, aphidicolin, in cases where there were negative or equivocal findings in the first RS comet experiment.

Laboratories from Europe and the USA participated in the testing of 30 blinded chemicals. After decoding, independent statistical analysis revealed an overall accuracy (concordance) of 80 % (sensitivity 73 %, specificity 87 %) when compared to *in vivo* animal genotoxicity test outcomes. After the IWGT meeting, further blinded chemicals were tested to increase the overlap with chemicals tested in the RSMN assay. This resulted in an increase of the RS comet assay's predictivity, to an overall accuracy of 83 % (sensitivity 77 %, specificity 88 %) [17]. The positive predictivity increased when the RS comet and RSMN assays are combined in a test battery approach [10].

In addition, the suggested use of the RS comet as a 2nd tier assay, specifically as a follow-up to positive findings from the bacterial reverse mutation test (OECD 471 [19]), was discussed during the Tokyo IWGT meeting. Cases were reported, in which data on three hair dyes were submitted to the Scientific Committee on Consumer Safety (SCCS), an independent expert committee of the EU Commission that provides opinions on health and safety risks of non-food consumer products, including cosmetic ingredients. Negative data obtained with the RS comet assay were accepted as part of a weight-of-evidence approach, and the hair dyes were considered "safe for use" based on all available data (SCCS/1531/14 [20], SCCS/1563/15 [21], SCCS/1572/16 [22]). Since 2014, the SCCS has recommended using both the RS comet and RSMN assays as a follow-up for suspected misleading positive results from the standard *in vitro* test battery (SCCS/1532/14 [23]). Importantly, the *in vivo* comet assay efficiently detected *in vivo* and *in vitro* mutagens [24,25] and the comet endpoint was therefore considered appropriate for follow-up testing of mutagenic substances in the *in vitro* battery.

2.3.2. The reconstructed skin micronucleus test (RSMN): validation data

The multi-year validation efforts for the RSMN using the EpiDerm™skin model have been finalized and the chemicals used for validation were decoded shortly before the Tokyo IWGT meeting. Detailed experimental data, as well as a preliminary analysis of the validation outcomes, were presented to the WG (for details see Pfuhler et al. [18]). Assay development was briefly described [26], transferability and protocol optimization efforts [27] were highlighted, and assay validity and data evaluation criteria were presented (for details see Pfuhler et al. [18]). The outcome of the blinded testing of over 40 coded chemicals was presented, showing an overall assay accuracy of 84 %, with a sensitivity of 80 % and a specificity of 87 % when compared to *in vivo* genotoxicity outcomes. The WG concluded that the RSMN assay is an acceptable alternative to the *in vivo* test and that the high predictivity also demonstrates that the test complies with all requirements to be accepted as a 2nd tier test. It should be noted that the final validation outcome as per independent analysis of a biostatistician [18] differs slightly from the above numbers presented to the WG in November 2017.

The use of the RSMN in the risk assessment process was discussed via a case study on a chemical that tested positive in one or both of the standard genotoxicity assays (i.e., mutagenicity and clastogenicity/aneugenicity *in vitro*). Para-phenylene diamine (PPD) tested positive in the Ames test but negative in the *Hprt* mammalian cell mutagenicity assay, and positive for clastogenicity *in vitro*; PDD, however, had a negative *in vivo* genotoxicity profile (SCCS/1443/11 [28]). Its apparent metabolic detoxification in the skin was supported by data indicating that it is N-acetylated when applied to human volunteers in a hair dye formulation [29]), and N-acetylated PPD was completely nongenotoxic *in vitro,* including in the Ames test and the *in vitro* comet assay [30]. Thus, PPD appeared to be a good candidate for a case study for follow-up of an *in vitro* genetox positive test article with skin-based 2nd tier models. Subsequent testing in two independent experiments showed negative results for PPD in the RSMN test (Fig. 4), indicating that the skin-based assay could have been used as an alternative to the animal studies to demonstrate this chemical's lack of genotoxicity via the dermal exposure route.

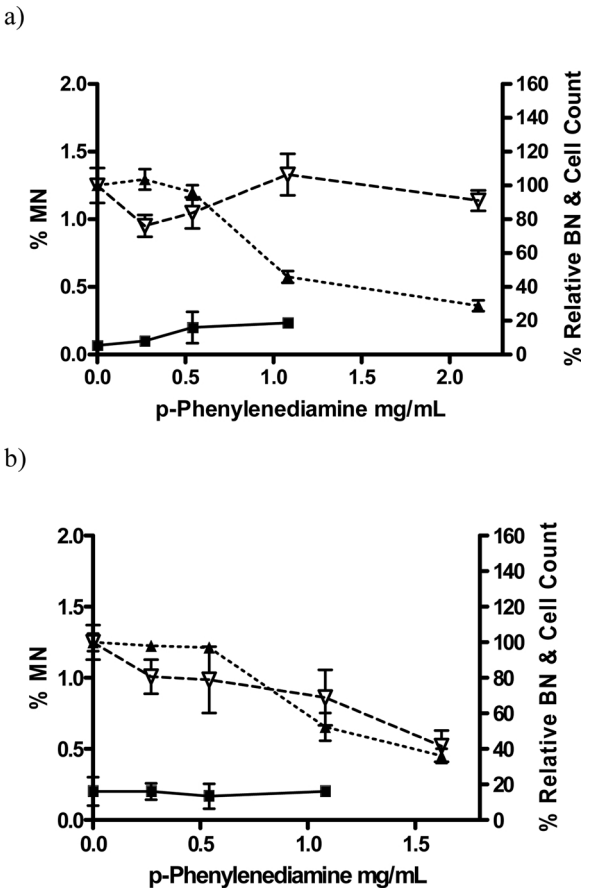

Fig. 4. a, b: RSMN assay with p-phenylene diamine – two independent experiments.

2.3.3. Establishment of a GLP method and proof of performance

Following publication of the CE standard method by Dahl et al. in 2011 [27], Covance Laboratories performed an in-house validation of the 3D RSMN assay independently of the work described in Section 2.3.2 above in order to establish a GLP protocol for testing unknown test substances. A summary of the results obtained from nine chemicals using the "standard method" 48-h exposure protocol was presented to the WG (Table 1). In addition, mixed results exist in the public literature for two metabolically activated genotoxins, cyclophosphamide and BaP, so new data generated following the 72-h exposure protocol (as described by Aardema et al. [26]) was also presented. This modified protocol incorporates a third addition of the test substance at the 48-h timepoint, followed by harvesting the cells at the 72-h timepoint.

Table 1
Summary of the testing (using EpiDerm™) for the standard 48 -h method as performed at Covance Laboratories when compared to the available data in the literature for nine chemicals. From Kidd [31], by permission of Oxford University Press.

Genotoxicity	Compound	Covance 48-hour	Published Literature 48-hour
Direct Acting clastogens	Mitomycin C	Positive	Positive
	Methyl methansulphonate	Positive	Positive
	n-Ethylnitrosourea	Positive	Positive
Aneugens	Vinblastine	Positive	Positive
Metabolically activated clastogens	Cyclophosphamide	Negative	Mixed results
	Benzo(a)pyrene	Negative	Mixed results
Non-Genotoxins	4-Nitrophenol	Negative	Negative
	Cyclohexanone	Negative	Negative
	2-Ethyl-1,3-hexanediol	Negative	Negative

Table 2
Statistical Power calculations – 1000 cells scored for 2 replicates per group. From Kidd [31] with permission.

Control Incidence (Mean MNBN cell frequency (%))	Test Group(1) (Mean MNBN cell frequency (%))	Difference (Test – Control) (Mean MNBN cell frequency (%))	Fold difference (Test / Control)
0/2000 (0.000)	5/2000 (0.250)	0.250	N/A
1/2000 (0.050)	7/2000 (0.350)	0.300	7.00
2/2000 (0.100)	9/2000 (0.450)	0.350	4.50
3/2000 (0.150)	10/2000 (0.500)	0.350	3.33
4/2000 (0.200)	12/2000 (0.600)	0.400	3.00
5/2000 (0.250)	13/2000 (0.650)	0.400	2.60
6/2000 (0.300)	15/2000 (0.700)	0.450	2.50

(1) Response in Test group that could be detected as statistically significant. Assuming a control incidence of 0.100 %, 1000 cells scored for 2 replicates would allow a difference of approximately 0.350 % (4.50 fold increase) between groups to be detected as statistically significant at the 5 % level using a one-sided Chi-squared test.

Replication Index (RI) data for the solvent controls between the 48- and 72-h exposure protocols indicated an increase in the number of cells undergoing division during the longer (72-h) exposure. In addition, the positive control data (for mitomycin C and vinblastine) also showed increases in the percentage of micronucleated binucleate cells (% MNBN) when using the 72-h exposure protocol compared to the responses observed using the standard 48-h method. Finally, theoretical and experimental data (Tables 2 and 3) were presented on the impact of scoring 1000 binucleate (BN) cells from each of two or three treated tissues per test concentration [31]. Both the theoretical and practical data showed that scoring 1000 binucleate cells from each of three tissues: 1) reduced the test substance concentration where an increase in MNBN cells was statistically significant (compared to the concurrent negative control) and 2) reduced the occurrence of scoring zero MNBN per 1000 BN scored. The RI data, % MNBN data and results varying by the number of cells scored were discussed, with the detailed experimental results and supporting data now published [31]. Documentation of training, management approval of SOPs and subsequent implementation, as well as independent Quality Assurance (QA) review of the resultant SOPs and protocol are GLP requirements and were satisfied as part of the laboratory's validation process.

Table 3
Statistical Power calculations - 1000 cells scored for 3 replicates per group. From Kidd [31], with permission.

Control Incidence (Mean MNBN cell frequency (%))	Test Group(1) (Mean MNBN cell frequency (%))	Difference (Test – Control) (Mean MNBN cell frequency (%))	Fold difference (Test / Control)
0/3000 (0.000)	5/3000 (0.167)	0.167	N/A
1/3000 (0.033)	7/3000 (0.233)	0.200	7.00
2/3000 (0.067)	9/3000 (0.300)	0.233	4.50
3/3000 (0.100)	10/3000 (0.333)	0.233	3.33
4/3000 (0.133)	12/3000 (0.400)	0.267	3.00
5/3000 (0.167)	13/3000 (0.433)	0.267	2.60
6/3000 (0.200)	15/3000 (0.500)	0.300	2.50

(1) Response in Test group that could be detected as statistically significant. Assuming a control incidence of 0.100 %, 1000 cells scored for 3 replicates would allow a difference of approximately 0.233 % (3.33 fold increase) between groups to be detected as statistically significant at the 5 % level using a one-sided Chi-squared test.

2.3.4. Method transfer to China and initial validation data

The RSMN protocol used for the CE validation also was adapted to the Episkin™ model with minor changes, i.e. milder conditions for cell dissociation and fixation onto the slides, to accommodate differences between the different skin models. A description of the protocol and the experiments that were performed are available in the paper by Chen et al. [32]. The initiative started at L'Oréal R&I China who then transferred the method to two other Chinese laboratories. A validation effort is currently ongoing that uses the Episkin™ model and that is supported scientifically by the CE genotoxicity taskforce. All three laboratories have shown proficiency in performing the RSMN assay. At the time of the IWGT workshop in Tokyo, results for 5 compounds tested in two of the participating laboratories (Guangdong Center for Disease Control and L'Oréal R&I China), and 2 compounds from the third laboratory (Zheijiang IFDC) were presented. Based on the 5 compounds, a 100 % inter-laboratory reproducibility was observed (see Table 4). One compound with mixed results *in vivo* (cyclohexanone) was found to be negative by the two laboratories that tested it. Due to the inconsistency of the *in vivo* results for this compound, it was dropped from the list of compounds for the validation. It also was emphasized that the validation of an additional skin model for the RSMN that is produced in China would be advantageous since shipment and export/import of these tissues can be difficult and laborious. This would also enable broader access to and acceptance of the RSMN in Asia.

2.4. Extension of applicability domain – example: nanomaterials

The liver, airway and skin 3D models described above have all been evaluated for their use in the genotoxicity testing of engineered nanomaterials. Iron oxide nanoparticles (both maghemite and magnetite) were used to evaluate the performance of the CBMN assay using 3D HepG2 liver spheroids. Indeed, the 3D liver spheroids were capable of detecting chromosomal damage induced by the oxidative species

Table 4
Outcome of the pre-validation testing efforts from 3 Chines laboratories.

Chemical name	CAS No.	*In vivo* micronucleus outcome	RSMN outcome *L'Oréal*	RSMN outcome *GDCDC*	RSMN outcome *ZIFDC*
			48h	48h	48h
Mytomycin C	50-07-7	Positive	Positive	Positive	Positive
Vinblastine	143-67-9	Positive	Positive	Positive	Positive
5-Fluorouracil	120-83-2	Positive	Positive	Positive	Not tested
2,4-Dichlorophenol	108-94-1	Negative	Negative	Negative	Not tested
Cyclohexanone	108-94-1	Negative/Positive	Negative	Negative	Not tested

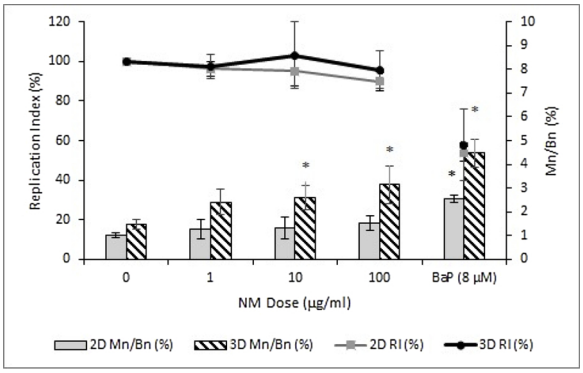

Fig. 5. Micronucleus frequency induced by dextran coated iron oxide nanoparticles in HepG2 cells cultured in standard 2D monolayer format and in 3D spheroid format. BaP at 8 µM was included as a positive control. N = 3; * p ≥ 0.05.

typically associated with these nanomaterials [33] (Fig. 5). A challenge with 3D models is the ability to evaluate the penetration of nanoparticles across the biological barrier presented by the 3D liver spheroid. This was achieved by the use of x-ray fluorescence mapping, which demonstrated that the iron oxide nanoparticles were primarily concentrated within the first few cell layers of the spheroid. Some nanoparticles, however, were able to penetrate up to 50 µm into the spheroid, although no material was evident in the spheroid core. Transmission electron microscopy (TEM) performed on the outer cell layers of the spheroid structure demonstrated that the nanoparticles were internalized by these cells.

Standard genotoxicity tests are typically based on a single cell type; however, a key observation in *in vivo* genotoxicity studies conducted with nanomaterials is that the DNA damage induced is often the result of secondary genotoxicity associated with inflammation [34,35]. Lung co-culture models comprised of both human lung epithelial cells and differentiated macrophages have therefore been established and used for evaluating genotoxicity using the CBMN assay. Dextran-coated ultrafine superparamagnetic iron oxide nanoparticles (USPION) were tested in this lung co-culture model, which demonstrated that dFe_3O_4 nanoparticles illicit a genotoxic response in the co-culture, which is absent when tested in lung epithelial cells only (*i.e.* standard 2D CBMN assay; Fig. 6) [36]. When cell uptake was evaluated in the co-culture model, the nanoparticles were only located in macrophages and were not internalized in the lung epithelial cells (where the genotoxicity was detected), thereby indicating a secondary genotoxic response was evident in this advanced culture system [36]. This particular model utilized a co-culture system, employing a mixture of two cell types, overlaid upon each other. This study therefore demonstrates a potential advantage of advanced models comprised of multiple cell types in detecting mechanisms *in vitro* which have not previously been detected in standard genotoxicity testing systems.

The 3D RSMN assay also has been used for the genotoxicity evaluation of silica nanoparticles of two different sizes, 16 and 85 nm diameter [37]. The ability of these materials to induce chromosomal damage was compared across both the 3D RSMN and the standard 2D

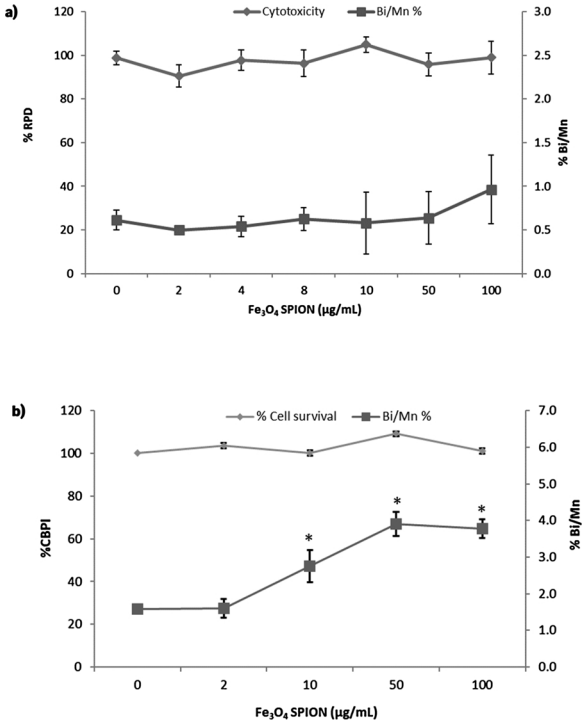

Fig. 6. a, b: Quantification of chromosomal damage and cell viability of 16HBE14o- cells following dSPION exposure. (a) Mono-cultured 16HBE14o- cells treated with Fe_3O_4; (b) lung co-culture model consisting of 16HBE14o- lung epithelial cells and differentiated THP-1 macrophages, treated with Fe_3O_4 dSPION. *p < 0.05 when compared to negative control (0 µg/ml). MMC (0.1 µg/ml) was used as positive control (micronuclei frequency 4.01 %).
Fig. 6 was reproduced with permission from Evans et al., 2019 [35]; this open access article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), and the article was published by BMC (part of Springer Nature).

MN assay after standardising the test substance concentrations in the 2D and 3D assays by determining the total nanoparticle mass to cell number tested. In this study, all exposures ≥100 µg/mL in the 2D CBMN assay resulted in significant increases in MN frequency as both test materials were able to readily enter the cells (as determined by TEM evaluation). In contrast, 2D-equivalent exposures to the 3D models caused no significant DNA damage. Uptake analysis in the 3D skin models using TEM revealed the nanomaterials that had been applied topically were not capable of penetrating the 3D microarchitecture of the tissue model; thus, there was no direct exposure of the test nanoparticles to the living cells in the model [37]. This study demonstrated that, when the RSMN protocol is applied to testing nanomaterials, it is important to ensure that the nanomaterials are not used at an excessively high concentration, which would compromise the surface of the skin construct and lead to an artifactual result. One such extreme concentration of silica nanoparticles was tested by Wills and colleagues, who noted that, although overt toxicity was recorded at this concentration, it was likely artifactual. Surface imaging of the 3D RS tissue

at this toxic concentration revealed excessive particle coverage that likely would have inhibited gas exchange at the air interface. Nonetheless, this study demonstrated that the use of RS models for nanomaterials offers a more realistic biological barrier that better represents the protective nature of the skin's 3D cellular microarchitecture, thereby improving the physiological relevance of the genotoxicity testing results.

3. Discussion

3.1. Why use 3D tissues for genotoxicity testing, and how do they fit into a testing strategy?

The WG initially discussed the status of development for RS, liver, and airway 3D tissue-based genotoxicity assays, as well as their fit within a genotoxicity testing strategy. In concordance with a previous expert review [5] and based on the data presented, the IWGT WG agreed that *'3D tissue models offer a more 'in-vivo-like' behavior for key parameters like cell viability, proliferation, differentiation, morphology, gene and protein expression, and function and therefore provide a valuable complement to the classical '2D' cell culture-based assays'*. Using more complex, *in-vivo*-like systems is a strong trend in toxicology and in pharmacology that has gained much attention in recent years and, as endorsed by this WG for tissue-based genotoxicity assays, is generally seen to provide data that are more relevant to evaluating genotoxicity in humans than traditional '2D' assays, [2,3,38,39].

With regard to their genotoxicity testing strategy fit, the WG initially discussed whether the 3D assays could be used as 1st tier assays to replace the standard *in vitro* 2D testing battery. This was not seen currently as a preferred option since it is assumed that regulatory agencies would like to see a mutational endpoint, e.g. gene mutations, as part of a submission. It also was pointed out that starting a genotoxicity testing program with the 3D assays would be a more time consuming and expensive approach compared to using them as follow-up for positive results from the standard 2D test battery. With tissue models being more complex, 3D assays are technically more difficult to perform and, consequently, their throughput is reduced when compared to 2D assays. One additional argument against using 3D assays as 1st tier tests was that the best use of 3D tissues would be to choose the respective model according to exposure conditions, e.g., RS tissues for dermal exposures, airway tissues for inhalation exposures, and liver tissues for systemic/oral exposures. The WG agreed that such a concept '*allows for exposure-specific investigations'*. It should also be noted that most 3D models are based on human cells, enhancing the relevance of the results to assessing potential human genotoxicity. To summarize the discussion relating to strategic fit, the WG agreed that '*3D tissue-based genotoxicity assays can be used as 2nd tier assays to follow-up on positive results from standard in vitro assays*' and sees the strength of 3D genotoxicity assays as an alternative to animal studies.

Subsequently, the WG discussed the strategic use of these assays in a test battery approach. In this context, the expectation was that the assays would be used as follow-up of positive results from the 1st tier *in vitro* testing battery (e.g., Ames and *in vitro* MN test). The WG agreed that the use of an assay (consisting of a specific tissue model and a genotoxicity read-out) for regulatory purposes will depend on its validation status. In order to be considered useful as 2nd tier assays '*it is important that, for each tissue model, the full range of genotoxic damage (leading to mutagenicity, clastogenicity, aneugenicity) can be detected'*. Therefore, while it would be ideal that each 3D tissue type would be validated for all genotoxicity endpoints at the same time, this may not always be the case. It was therefore emphasized that '*these assays can be used in combination or alone depending on the scenario since these 3D models are suitable for endpoint-specific follow-up of positives from standard in vitro testing battery'*. With respect to the positioning of 3D assays, the following three scenarios were discussed and agreed:

- in vitro clastogenicity/aneugenicity positives lead to 3D MN testing,
- in vitro mutagenicity positives lead to 3D comet
- in vitro mutagenicity & clastogenicity/aneugenicity positives lead to both 3D comet and 3D MN.

In this context, it was emphasized that the more frequent scenario likely will address a single positive from the standard battery. It is recognized by all WG members that there are other assays/readout methods available that could potentially be used in 3D organoid models (e.g., a gene mutation readout such as *Pig*-a) [40] and could serve as a follow-up of positive results from the standard *in vitro* battery. The focus on the comet and MN assays was triggered by the existing experimental datasets for the 3D skin, airway and liver tissues, which was limited to these endpoints.

3.2. How can advantages of in-vivo-like '3D' models be leveraged further?

Most of the experimental data presented at the meeting were generated with chemicals; however, the WG acknowledged that the specific advantages of 3D models in terms of more *in-vivo*-like ADME properties, in particular penetration/absorption, make them an interesting model for studying potential genotoxic effects of nanomaterials. Examples presented included the determination of the genotoxic potential of silica nanomaterials after dermal exposure. While giving a positive response in standard 2D cell culture systems, RS models offer a more realistic biological barrier for testing effects via the dermal exposure route and the lack of a genotoxic response in the RS models likely reflect the *in vivo* situation since silica nanomaterials do not penetrate the dermis *in vivo* [37]. Conversely, in a co-culture model with macrophage and lung epithelial cells, it was shown that dFe_3O_4 nanoparticles were taken up only by macrophage cells and not by the lung epithelial cells. Genotoxic effects, likely secondary in nature, were detected in lung epithelial cells in the 3D co-culture system but not when exposed in 2D lung cell cultures where no macrophages were present [36]. These examples demonstrate advantages of such complex *in vitro* models, leading the WG to conclude that '*3D tissue-based assays provide a more realistic test system to study particulate materials (e.g. nanomaterials), compared to 2D test systems'*. It was also emphasised that it is critical to have a clear understanding of the fate of the nanomaterials used under the respective assay conditions [41] and also to include measurements of cellular uptake (e.g. electron microscope analysis of penetration / uptake) to help understand the outcome of the assay. *In situ* characterization and uptake measurements have been recommended by several expert committees and are already included in a genotoxicity testing guidance for nanomaterials [41]. The WG also recommends that any future guideline for '3D' model-based genotoxicity assays consider these factors when testing nanomaterials.

3.3. Discussion of the validation status of 3D liver, airway and skin-based genotoxicity assays

The three different tissue models are at various stages of development and validation, with the RS model being the only one that has undergone formal validation. For assays based on airway tissue and liver organoids, existing data are sparse [8,42–44]. While consensus statements specific to each tissue model are captured further below, for both the liver and airway the WG encourages '*development of a robust protocol and testing of an initial set of chemicals representing expected positives and negatives covering different chemical classes before following up with other validation steps, like transferability, intra- and inter-laboratory reproducibility, applicability domain and predictive capacity* [45–47]*'*. This supports the goal of having validated methods that enable exposure-specific assessments. The respective models have the capacity to represent dermal, oral and inhalation exposure routes in safety assessment and there was agreement supporting efforts directed at the continued development and validation of these models.

Next, the organ-specific status was discussed in more detail and the WG agreed that for 3D liver spheroids *'Initial data show that the MN assay can be applied to 3D liver spheroids and the WG encourages further development of this assay'*. During the discussion of the protocol, it was recommended that for further protocol development, exposure time and the concentration of cytochalasin B could be further optimized, and the determination of cell cycle time might be useful in this context. Experience with endpoints other than MN, however, is much more limited and it was concluded that *'the lack of 3D liver assays that can detect substances that induce gene mutation is considered a gap and the development of such an assay is strongly encouraged'*. Some preliminary data indicate that the comet assay can be useful in this respect [48]. It was also mentioned that HepaRG cell-derived 3D liver spheroids are being investigated in detail within the framework of the EU HORIZON 2020 project PATROLS (www.patrols-h2020.eu), including characterization of their metabolic competence.

For 3D airway model-based assays, the WG concluded, based on the data presented, that *'Initial data show that the comet assay can be applied to the 3D airway models and the WG encourages further development of this assay'*. It was emphasized that *'the lack of 3D airway assays that can detect aneugenicity is considered a gap and the development of such an assay is strongly encouraged'*. In this context, several WG members pointed out that the limited proliferation rate of the cells in the 3D airway model currently makes the MN assay problematic. More effort is needed to resolve this problem. User experience was shared, and it was pointed out that proliferation rates will depend on the cell type and source used for the model. It was also suggested that, instead of MN, testing of ɣH2AXmay be helpful, since cells do not need to divide for this readout system. However, this is an indicator test that does not detect aneugens, and further research in this area is therefore highly encouraged.

It also was pointed out that only limited information on the metabolic competence of the 3D lung models was available at the time of the WG meeting [49] and that a better understanding in this regard would be important. It is encouraging that, subsequent to the IWGT WG meeting, data on the metabolic competence of the human ALI airway model has begun to appear [8].

In contrast to liver and airway 3D models, development of 3D skin model-based genotoxicity assays started over a decade ago and they have been explored for their utility as follow-up assays for dermal exposures. This work is in accordance with the recommendation of Maurici and colleagues that was developed at an expert meeting charged with the question of what *in-vitro*-only approaches would be most appropriate for genotoxicity hazard identification [50]. This question originated in the context of the EU Directive for cosmetic ingredients which prohibits animal testing for genotoxicity [12] and the first experimental data were published shortly after by Curren et al. [51]. Based on the data presented at the Tokyo IWGT meeting, which reflect the outcome of a multi-year international validation exercise [17,18], the WG concluded that '*following extensive validation and practice (in use for over 10 years) we are now in the position to define standard protocols for the 3D skin comet and RSMN assays'*. Furthermore, it was concluded that *'transferability of the assays to a large number of laboratories across 3 continents has been demonstrated',* and it was emphasized that *'the assays are now available at several Contract Research Organisations (CRO) and are performed under GLP (Good Laboratory Practice)'*.

There was considerable discussion specific to the RSMN protocol, since initial work for this assay used a 48 h treatment protocol (two exposures with a 24 h interval), whereas there is growing evidence from the validation dataset [18] as well as CRO experience [31] that a 72 h protocol with three exposures is superior since it increased the sensitivity of the assay. The WG concluded that '*validation data and CRO experience show that the 72 h protocol for the RSMN has higher sensitivity than the 48 h protocol and we therefore agree that the assay can be routinely performed using the 72 h protocol'*. Importantly, with this move to a 72 h protocol, the specificity of the assay decreased only marginally. It is expected that these findings will influence protocol modifications such that the initial testing of an unknown substance will start with a 72 h and 3-exposure protocol rather than with the original 48 h protocol.

The key elements of validation that help assess the maturity of an assay were then discussed, i.e. reproducibility, sensitivity, specificity and predictive capacity. It was concluded from the data presented that '*international validation studies with coded chemicals have demonstrated good intra- and inter-laboratory reproducibility of the methods',* and that *'through the testing of 56 coded chemicals across assays (3D skin comet and RSMN), the combination of the methods has been shown to be highly predictive of the expected genotoxicity in vivo'*. The WG members agreed that the RS comet and RSMN assays have reached an advanced stage of maturity and concluded *'that the 3D skin comet and micronucleus assays are now sufficiently validated to move towards the development of individual OECD Test Guidelines'*. Although the WG considered the assays to have undergone sufficient validation, it is envisaged that an independent peer review of the validation study will need to be carried out, as recommended in OECD Guidance Document 34 [45], before developing an OECD TG. This can be conducted by various organizations, including the OECD or specific "validation bodies" specializing in independent review of validation data, such as, for example, the EURL ECVAM Scientific Advisory Committee (ESAC).

4. Conclusions

The workshop demonstrated that extensive progress has been made in the last few years on the development of 3D genotoxicity models. Different 3D culture models are now available that represent the major routes of exposure: dermal, systemic (liver spheroids, organ on a chip), and inhalation. The IWGT WG believes that the 3D skin comet and MN assays are now sufficiently validated to begin the development of individual OECD Test Guidelines. Although the WG considered the assays to have undergone sufficient validation, an independent peer review of the validation study will need to be carried out as recommended by OECD Guidance Document 34.

Genotoxicity assays using the two other 3D models are at an early stage of development and still do not sufficiently cover the three key endpoints of genotoxicity. For the 3D airway model, clastogenicity and gene mutation can be measured, at least indirectly, by the comet assay, but the development of a 3D airway MN model for detection of aneuploidy is desirable. Likewise, for the 3D liver model, where MN can be detected, a test that can evaluate gene mutations is essential. In addition, for airway 3D models, '*more information on the metabolic competence of the cells is considered important.'* Finally, for both the liver and airway models the considerable hurdle of assay validation remains: thus, we *'recommend developing a robust protocol and testing of an initial set of chemicals representing expected positives and negatives covering different chemical classes before following up with other validation steps, being transferability, intra- and inter-laboratory reproducibility, applicability domain and predictive capacity'* [45,46].

Funding sources

The 3D skin comet and micronucleus validation outcome presented at IWGT was financially supported by the German Federal Ministry for Research and Technology (grant Nos. 0315226, 0316008) and by Cosmetics Europe, Belgium. The research on CeO_2 nanoparticles presented in chapter 2b was financially supported by funds from TNO, The Netherlands.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. Alepee, A. Bahinski, M. Daneshian, B. De Wever, E. Fritsche, A. Goldberg, J. Hansmann, T. Hartung, J. Haycock, H. Hogberg, L. Hoelting, J.M. Kelm, S. Kadereit, E. McVey, R. Landsiedel, M. Leist, M. Lubberstedt, F. Noor, C. Pellevoisin, D. Petersohn, U. Pfannenbecker, K. Reisinger, T. Ramirez, B. Rothen-Rutishauser, M. Schafer-Korting, K. Zeilinger, M.G. Zurich, State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology, Altex 31 (2014) 441–477.

[2] J. Augustyniak, A. Bertero, T. Coccini, D. Baderna, L. Buzanska, F. Caloni, Organoids are promising tools for species-specific in vitro toxicological studies, J. Appl. Toxicol. 39 (2019) 1610–1622.

[3] M. Weinhart, A. Hocke, S. Hippenstiel, J. Kurreck, S. Hedtrich, 3D organ models-Revolution in pharmacological research? Pharmacol. Res. 139 (2019) 446–451.

[4] S. Pfuhler, M. Fellows, J. van Benthem, R. Corvi, R. Curren, K. Dearfield, P. Fowler, R. Frotschl, A. Elhajouji, L. Le Hegarat, T. Kasamatsu, H. Kojima, G. Ouedraogo, A. Scott, G. Speit, In vitro genotoxicity test approaches with better predictivity: summary of an IWGT workshop, Mutat. Res. 723 (2011) 101–107.

[5] E. Zeiger, B. Gollapudi, M.J. Aardema, S. Auerbach, D. Boverhof, L. Custer, P. Dedon, M. Honma, S. Ishida, A.L. Kasinski, J.H. Kim, M.G. Manjanatha, J. Marlowe, S. Pfuhler, I. Pogribny, W. Slikker, L.F. Stankowski Jr., J.Y. Tanir, R. Tice, J. van Benthem, P. White, K.L. Witt, V. Thybaud, Opportunities to integrate new approaches in genetic toxicology: an ILSI-HESI workshop report, Environ. Mol. Mutagen. 56 (2015) 277–285.

[6] U.K. Shah, J.O. Mallia, N. Singh, K.E. Chapman, S.H. Doak, G.J.S. Jenkins, A three-dimensional in vitro HepG2 cells liver spheroid model for genotoxicity studies, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 825 (2018) 51–58.

[7] X. Cao, H. Lin, L. Muskhelishvili, J. Latendresse, P. Richter, R.H. Heflich, Tight junction disruption by cadmium in an in vitro human airway tissue model, Respir. Res. 16 (2015) 30.

[8] Q. Qin, Q. Wu, Y. Wang, R. Xiong, L. Guo, X. Fu, H. Rosenfeldt, M. Bryant, X. Cao, Effects of cellular differentiation in human primary bronchial epithelial cells: metabolism of 4-(methylnitrosamine)-1-(3-pyridyl)-1-butanone, Toxicol. In Vitro 55 (2019) 185–194.

[9] I.M. Kooter, M. Gröllers-Mulderij, M. Steenhof, D. Evert, F.A.A. van Acker, Y.C.M. Staal, P.C. Tromp, E. Schoen, C.F. Kuper, E. van Someren, Cellular effects in an in vitro human 3D cellular airway model and A549/BEAS-2B in vitro cell cultures following air exposure to cerium oxide particles at an air–Liquid interface, Appl. In Vitro Toxicol. 2 (2016) 56–66.

[10] S. Pfuhler, A. Kirst, M. Aardema, N. Banduhn, C. Goebel, D. Araki, M. Costabel-Farkas, E. Dufour, R. Fautz, J. Harvey, N.J. Hewitt, J. Hibatallah, P. Carmichael, M. Macfarlane, K. Reisinger, J. Rowland, F. Schellauf, A. Schepky, J. Scheel, A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity, A COLIPA analysis, Regul Toxicol Pharmacol 57 (2010) 315–324.

[11] S. Pfuhler, R. Fautz, G. Ouedraogo, A. Latil, J. Kenny, C. Moore, W. Diembeck, N.J. Hewitt, K. Reisinger, J. Barroso, The Cosmetics Europe strategy for animal-free genotoxicity testing: project status up-date, Toxicol. In Vitro 28 (2014) 18–23.

[12] EU, EU Regulation EC No. 1907/2006 OF the European Parliament and of the Councilof Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency, Amending Directive 1999/45/EC and Repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 As Well As Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. (2006).

[13] N.J. Hewitt, R.J. Edwards, E. Fritsche, C. Goebel, P. Aeby, J. Scheel, K. Reisinger, G. Ouedraogo, D. Duche, J. Eilstein, A. Latil, J. Kenny, C. Moore, J. Kuehnl, J. Barroso, R. Fautz, S. Pfuhler, Use of human in vitro skin models for accurate and ethical risk assessment: metabolic considerations, Toxicol. Sci. 133 (2013) 209–217.

[14] C. Wiegand, N.J. Hewitt, H.F. Merk, K. Reisinger, Dermal xenobiotic metabolism: a comparison between native human skin, four in vitro skin test systems and a liver system, Skin Pharmacol. Physiol. 27 (2014) 263–275.

[15] C. Jackh, V. Blatz, E. Fabian, K. Guth, B. van Ravenzwaay, K. Reisinger, R. Landsiedel, Characterization of enzyme activities of Cytochrome P450 enzymes, Flavin-dependent monooxygenases, N-acetyltransferases and UDP-glucuronyl-transferases in human reconstructed epidermis and full-thickness skin models, Toxicol. In Vitro 25 (2011) 1209–1214.

[16] K. Reisinger, V. Blatz, J. Brinkmann, T.R. Downs, A. Fischer, F. Henkler, S. Hoffmann, C. Krul, M. Liebsch, A. Luch, R. Pirow, A.A. Reus, M. Schulz, S. Pfuhler, Validation of the 3D Skin Comet assay using full thickness skin models: transferability and reproducibility, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 827 (2018) 27–41.

[17] S. Pfuhler, R. Pirow, T.R. Downs, A. Haase, N.J. Hewitt, S. Hoffmann, A. Luch, M. Merkel, C. Petrick, A. Said, M. Schäfer-Korting, K. Reisinger, Validation of the 3D Reconstructed Human Skin Comet Assay, an Animal-free Alternative for Following-up Positive Results From Standard in Vitro Genotoxicity Assays, Mutagenesis (2019) In press.

[18] S. Pfuhler, M. Aardema, R. S, J. Scheirer, R. Kulkarni, G. Mun, N. Wilt, E. Costin, R. Curren, B. Barnett, S. Hoffmann, N.J. Hewitt, Validation of the 3D Human Reconstructed Skin Micronucleus Assay (RSMN), an Animal-free Alternative for Following-up Positive Results From Standard in Vitro Genotoxicity Assays, Mutagenesis (2019) In preparation.

[19] OECD, Test No. 471: Bacterial Reverse Mutation Test, (1997).

[20] SCCS, Scientific Committee on Consumer Safety Opinion on Basic Brown 17 COLIPA N° B007, SCCS/1531/14, Revision of 18 June 2014, the SCCS Adopted This Opinion at Its 5th Plenary Meeting of 24 March, (2014).

[21] SCCS, Scientific Committee on Consumer Safety. Opinion on 2,6-Dihydroxyethylaminotoluene COLIPA N° A138, SCCS/1563/15, the SCCS Adopted This Opinion at Its 10th Plenary Meeting on 25 June, (2015).

[22] SCCS, Scientific Committee on Consumer Safety Opinion on N,N'-Bis-(2-hydroxyethyl)-2-nitro-p-phenylenediamine (B34), SCCS/1572/16, Final Version of 16 September 2016, the SCCS Adopted the Final Version of This Opinion by Written Procedure on 16 September, (2016).

[23] SCCS, Scientific Committee on Consumer Safety. Addendum to the SCCS's Notes of Guidance (NoG) for the Testing of Cosmetic Ingredients and Their Safety Evaluation, 8th Revision (SCCS/1501/12). This Addendum Replaces the Section 3-4.7 Mutagenicity/Genotoxicity and 3-4.8 Carcinogenicity of the NoG, From Pages 26 to 29. SCCS/1532/14, Revision of 22 October 2014. The SCCS Adopted this Addendum by written procedure on 9 April, (2014).

[24] D. Kirkland, D.D. Levy, M.J. LeBaron, M.J. Aardema, C. Beevers, J. Bhalli, G.R. Douglas, P.A. Escobar, C.S. Farabaugh, M. Guerard, G.E. Johnson, R. Kulkarni, F. Le Curieux, A.S. Long, J. Lott, D.P. Lovell, M. Luijten, F. Marchetti, J.J. Nicolette, S. Pfuhler, D.J. Roberts, L.F. Stankowski Jr., V. Thybaud, S.K. Weiner, A. Williams, K.L. Witt, R. Young, A comparison of transgenic rodent mutation and in vivo comet assay responses for 91 chemicals, Mutat. Res. Genet. Toxicol. Environ. Mutagen. 839 (2019) 21–35.

[25] D. Kirkland, Y. Uno, M. Luijten, C. Beevers, J. van Benthem, B. Burlinson, S. Dertinger, G.R. Douglas, S. Hamada, K. Horibata, D.P. Lovell, M. Manjanatha, H.-J. Martus, N. Mei, T. Morita, W. Ohyama, A. Williams, In vivo genotoxicity testing strategies: report from the 7th International Workshop on Genotoxicity Testing (IWGT), Mutat. Res. (2019) This issue.

[26] M.J. Aardema, B.B. Barnett, G.C. Mun, E.L. Dahl, R.D. Curren, N.J. Hewitt, S. Pfuhler, Evaluation of chemicals requiring metabolic activation in the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay, Mutat. Res. 750 (2013) 40–49.

[27] E.L. Dahl, R. Curren, B.C. Barnett, Z. Khambatta, K. Reisinger, G. Ouedraogo, B. Faquet, A.C. Ginestet, G. Mun, N.J. Hewitt, G. Carr, S. Pfuhler, M.J. Aardema, The reconstructed skin micronucleus assay (RSMN) in EpiDerm: detailed protocol and harmonized scoring atlas, Mutat. Res. 720 (2011) 42–52.

[28] SCCS, Scientific Committee on Consumer Safety Opinion on p-Phenylenediamine COLIPA N° A7, SCCS/1443/11, Revision of 18 September 2012, the SCCS Adopted This Opinion at Its 15th Plenary Meeting, of 26–27 June, (2012).

[29] G.J. Nohynek, J.A. Skare, W.J. Meuling, D.W. Hein, A.T. De Bie, H. Toutain, Urinary acetylated metabolites and N-acetyltransferase-2 genotype in human subjects treated with a para-phenylenediamine-containing oxidative hair dye, Food Chem. Toxicol. 42 (2004) 1885–1891.

[30] A. Zeller, S. Pfuhler, N-acetylation of three aromatic amine hair dye precursor molecules eliminates their genotoxic potential, Mutagenesis 29 (2014) 37–48.

[31] Kidd, Covance RSMN In-House Validation Data, Mutagenesis (2019), https://doi.org/10.1093/mutage.gez1037 in press.

[32] L. Chen, N. Li, Y. Liu, B. Faquet, N. Alépée, C. Ding, J. Eilstein, L. Zhong, Z. Peng, J. Ma, Z. Cai, G. Ouedraogo, RSMN on Episkin, Mutagenesis (2019) in press.

[33] N. Singh, G.J. Jenkins, B.C. Nelson, B.J. Marquis, T.G. Maffeis, A.P. Brown, P.M. Williams, C.J. Wright, S.H. Doak, The role of iron redox state in the genotoxicity of ultrafine superparamagnetic iron oxide nanoparticles, Biomaterials 33 (2012) 163–170.

[34] S. Pfuhler, T.R. Downs, A.J. Allemang, Y. Shan, M.E. Crosby, Weak silica nanomaterial-induced genotoxicity can be explained by indirect DNA damage as shown by the OGG1-modified comet assay and genomic analysis, Mutagenesis 32 (2017) 5–12.

[35] S.J. Evans, M.J. Clift, N. Singh, J. de Oliveira Mallia, M. Burgum, J.W. Wills, T.S. Wilkinson, G.J. Jenkins, S.H. Doak, Critical review of the current and future challenges associated with advanced in vitro systems towards the study of nanoparticle (secondary) genotoxicity, Mutagenesis 32 (2017) 233–241.

[36] S.J. Evans, M.J.D. Clift, N. Singh, J.W. Wills, N. Hondow, T.S. Wilkinson, M.J. Burgum, A.P. Brown, G.J. Jenkins, S.H. Doak, In vitro detection of in vitro secondary mechanisms of genotoxicity induced by engineered nanomaterials, Part. Fibre Toxicol. 16 (2019) 8.

[37] J.W. Wills, N. Hondow, A.D. Thomas, K.E. Chapman, D. Fish, T.G. Maffeis, M.W. Penny, R.A. Brown, G.J. Jenkins, A.P. Brown, P.A. White, S.H. Doak, Genetic toxicity assessment of engineered nanoparticles using a 3D in vitro skin model (EpiDerm), Part. Fibre Toxicol. 13 (2016) 50.

[38] S.C. Ramaiahgari, S.S. Ferguson, Organotypic 3D HepaRG liver model for assessment of drug-induced cholestasis, Methods Mol. Biol. 1981 (2019) 313–323.

[39] L.M. Norona, D.G. Nguyen, D.A. Gerber, S.C. Presnell, E.L. LeCluyse, E. Highlight, Modeling compound-induced fibrogenesis in vitro using three-dimensional bioprinted human liver tissues, Toxicol. Sci. 154 (2016) 354–367.

[40] B. Kopp, J. Vignard, G. Mirey, V. Fessard, D. Zalko, L. Le Hgarat, M. Audebert, Genotoxicity and mutagenicity assessment of food contaminant mixtures present in the French diet, Environ. Mol. Mutagen. 59 (2018) 742–754.

[41] R. Elespuru, S. Pfuhler, M.J. Aardema, T. Chen, S.H. Doak, A. Doherty, C.S. Farabaugh, J. Kenny, M. Manjanatha, B. Mahadevan, M.M. Moore, G. Ouedraogo, L.F. Stankowski Jr., J.Y. Tanir, Genotoxicity assessment of nanomaterials: recommendations on best practices, assays, and methods, Toxicol. Sci. 164 (2018) 391–416.

[42] A. Reus, Y. Staal, J. van Triel, F. van Acker, F. Kuper, Human 3D airway models to explore in vivo inhalation, Eur. Respir. J. 38 (2011) p3093.

[43] C. Frieke Kuper, M. Grollers-Mulderij, T. Maarschalkerweerd, N.M. Meulendijks, A. Reus, F. van Acker, E.K. Zondervan-van den Beuken, M.E. Wouters, S. Bijlsma, I.M. Kooter, Toxicity assessment of aggregated/agglomerated cerium oxide

nanoparticles in an in vitro 3D airway model: the influence of mucociliary clearance, Toxicol. In Vitro 29 (2015) 389–397.
[44] A. Haase, N. Dommershausen, M. Schulz, R. Landsiedel, P. Reichardt, B.C. Krause, J. Tentschert, A. Luch, Genotoxicity testing of different surface-functionalized SiO2, ZrO2 and silver nanomaterials in 3D human bronchial models, Arch. Toxicol. 91 (2017) 3991–4007.
[45] OECD, OECD Series on Testing and Assessment. Number 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazary Assessment. ENV/JM/MONO(2005)14, 18-Aug, (2005).
[46] T. Hartung, S. Bremer, S. Casati, S. Coecke, R. Corvi, S. Fortaner, L. Gribaldo, M. Halder, S. Hoffmann, A.J. Roi, P. Prieto, E. Sabbioni, L. Scott, A. Worth, V. Zuang, A modular approach to the ECVAM principles on test validity, Altern. Lab. Anim. 32 (2004) 467–472.
[47] R. Curren, J.A. Southee, H. Spielmann, M. Liebsch, J. Fentem, M. Balls, The role of prevalidation in the development, validation and acceptance of alternative methods. ECVAM Prevalidation Task Force Report 1, ATLA 23 (1995) 409–470.
[48] E. Elje, M. Hesler, E. Runden-Pran, P. Mann, E. Mariussen, S. Wagner, M. Dusinska, Y. Kohl, The comet assay applied to HepG2 liver spheroids, Mutat. Res. 845 (2019) 403033.
[49] J. Boei, S. Vermeulen, B. Klein, P.S. Hiemstra, R.M. Verhoosel, D.G.J. Jennen, A. Lahoz, H. Gmuender, H. Vrieling, Xenobiotic metabolism in differentiated human bronchial epithelial cells, Arch. Toxicol. 91 (2017) 2093–2105.
[50] D. Maurici, M. Aardema, R. Corvi, M. Kleber, C. Krul, C. Laurent, N. Loprieno, M. Pasanen, S. Pfuhler, B. Phillips, E. Sabbioni, T. Sanner, P. Vanparys, Alternative (non-animal) methods for cosmetics testing: current status and future prospects. Chapter 3.7: Genotoxicity/Mutagenicity, Altern. Lab. Anim. 33 (Suppl. 1) (2005) 117–130.
[51] R.D. Curren, G.C. Mun, D.P. Gibson, M.J. Aardema, Development of a method for assessing micronucleus induction in a 3D human skin model (EpiDerm), Mutat. Res. 607 (2006) 192–204.