

Original Manuscript

Validation of the 3D reconstructed human skin micronucleus (RSMN) assay: an animal-free alternative for following-up positive results from standard *in vitro* genotoxicity assays

Stefan Pfuhler^{1,*}, Thomas R. Downs¹, Nicola J. Hewitt², Sebastian Hoffmann³, Greg C. Mun⁴, Gladys Ouedraogo⁵, Shambhu Roy⁶, Rodger D. Curren⁴ and Marilyn J. Aardema⁷

¹Procter & Gamble Co., 8700 Mason-Montgomery Road, Mason, OH 45040, USA, ²Cosmetics Europe, Avenue Herrmann-Debroux 40, B-1160 Brussels, Belgium, ³seh consulting + services, Stembergring 15, 33106 Paderborn, Germany, ⁴Institute for In Vitro Sciences, Inc., 30 West Watkins Mill Road, Gaithersburg, MD 20878, USA, ⁵L'Oreal Research & Innovation, 1 Av Eugene Schueller, 93600 Aulnay-sous-Bois, France, ⁶MilliporeSigma, 14920 Broschart Road, Rockville, MD 20850, USA, ⁷Marilyn Aardema Consulting LLC, 5315 Oakbrook Dr., Fairfield, OH 45014, USA

*To whom correspondence should be addressed. Tel: +1 513 622 1163; E-mail: pfuhler.s@pg.com

Received 15 October 2020; Editorial decision 19 November 2020; Accepted 25 November 2020.

Abstract

In vitro test batteries have become the standard approach to determine the genotoxic potential of substances of interest across industry sectors. While useful for hazard identification, standard *in vitro* genotoxicity assays in 2D cell cultures have limited capability to predict *in vivo* outcomes and may trigger unnecessary follow-up animal studies or the loss of promising substances where animal tests are prohibited or not desired. To address this problem, a team of regulatory, academia and industry scientists was established to develop and validate 3D *in vitro* human skin-based genotoxicity assays for use in testing substances with primarily topical exposure. Validation of the reconstructed human skin micronucleus (RSMN) assay in MatTek Epi-200™ skin models involved testing 43 coded chemicals selected by independent experts, in four US/European laboratories. The results were analysed by an independent statistician according to predefined criteria. The RSMN assay showed a reproducibly low background micronucleus frequency and exhibited sufficient capacity to metabolise pro-mutagens. The overall RSMN accuracy when compared to *in vivo* genotoxicity outcomes was 80%, with a sensitivity of 75% and a specificity of 84%, and the between- and within-laboratory reproducibility was 77 and 84%, respectively. A protocol involving a 72-h exposure showed increased sensitivity in detecting true positive chemicals compared to a 48-h exposure. An analysis of a test strategy using the RSMN assay as a follow-up test for substances positive in standard *in vitro* clastogenicity/aneugenicity assays and a reconstructed skin Comet assay for substances with positive results in standard gene mutation assays results in a sensitivity of 89%. Based on these results, the RSMN assay is considered sufficiently validated to establish it as a 'tier 2' assay for dermally exposed compounds and was recently accepted into the OECD's test guideline development program.

Introduction

The genotoxicity hazard of a substance is generally explored by a battery of *in vitro* assays covering gene mutations, as well as structural and numerical chromosomal aberrations. Such a standard *in vitro* test battery will typically consist of at least two assays, often including the Ames assay and the *in vitro* micronucleus (MN) or chromosomal aberration assay, since no single assay can detect all genotoxicity endpoints. In some jurisdictions, showing the absence of a genotoxicity hazard *in vitro* is sufficient to classify a substance as non-genotoxic (e.g. REACH, EU Cosmetics Directive), while others (e.g. ICH guidelines for pharmaceuticals) always require testing in animals. As described previously in more detail by Pfuhler et al. (1), for all jurisdictions, positive results from the standard *in vitro* battery were traditionally investigated further by follow-up testing using animals. However, a paradigm shift away from animal studies is well under way, driven by animal welfare concerns and by questions about human relevance of rodent toxicity data. For example, the 7th Amendment to the Cosmetics Directive of the European Commission has prohibited the use of *in vivo* genotoxicity tests for cosmetic ingredients since 2009 (2), which means that a positive result from an *in vitro* standard genotoxicity assay would result in the loss of this ingredient from further development. Where *in vivo* testing is still allowed, a 'misleading positive' (MP) finding would trigger potentially unnecessary *in vivo* follow-up testing and animal usage, and it, therefore, seems desirable to design and validate higher tier *in vitro* assays that can serve the purpose of verifying results from the standard *in vitro* test battery.

There are currently no validated *in vitro* assays to address potential genotoxicity via the dermal exposure route. To close this gap, Cosmetics Europe initiated and led the development and validation of skin-based assays that combine classical genotoxicity endpoints with existing 3D skin models utilising reconstructed human skin (RS) already in use for other endpoints (3–6). 3D skin tissue constructs are logical follow-up tools for standard genotoxicity assays that utilise '2D' cell systems for multiple reasons, including the fact that the 3D skin tissue models are composed of primary cells of human origin (7,8). This feature was suggested to be important to avoid the generation of MP results during a workshop hosted by the European Union Reference Laboratory for alternatives to animal testing (EURL-ECVAM) (9). In addition, the use of primary human skin cells not only removes concerns due to rodent species-specific effects, it also eliminates the need for external metabolising systems such as rat liver S9. This is because the 3D skin models exhibit xenobiotic metabolism capabilities similar to native human skin, which therefore better reflects metabolism occurring by a dermal route of exposure (10–13). These efforts resulted in the development of the RS Comet assay (1,14,15) and the validation of the reconstructed human skin micronucleus (RSMN) assay (7,16,17). The MN endpoint provides a comprehensive basis for investigating chromosome damaging potential *in vitro* because both aneugens and clastogens can be detected (18) and is complementary to the Comet assay which can detect primary DNA damage leading to gene mutation (17). The combination of these assays is thought to cover all genotoxicity endpoints needed to follow-up on positive results from the standard 2D *in vitro* genotoxicity assays (9,17).

The RSMN assay in EpiDerm was developed early in the 2000s and prevalidation studies were conducted in mid-2000 (7,12,19). These early studies provided the foundation for the development of 3D skin models for genotoxicity testing. Proof-of-concept studies had shown that the cytokinesis-block methodology with cytochalasin B (CytoB) worked effectively in EpiDerm, resulting in a low and

reproducible background frequency of MN which enabled the successful detection of two model genotoxins (7). Prevalidation studies in the USA demonstrated intra- and inter-laboratory reproducibility in the RSMN assay in EpiDerm (12,19). Based on this, formal validation efforts by CosEU (formerly COLIPA) with involvement of ECVAM for the RSMN in EpiDerm started in 2007, the triggers and context of which are described in more detail by Pfuhler et al. (20). The validation program designed and funded by Cosmetics Europe was steered by its Genotoxicity Task Force, together with ECVAM and external experts involved to expand the assay to also include European laboratories (21). Two training workshops were held to standardise the protocol and harmonise the scoring of micronuclei, which were published in 2011 (16). The assay was transferred to two European laboratories and then three coded compounds were tested by three laboratories to demonstrate transferability and initial reproducibility within and between laboratories (21). Next, the number of chemicals tested in the RSMN was extended to include chemicals from more chemical classes and modes of action, which were tested in a double-blinded fashion. Initial results suggested a high predictive capacity (22).

For both the RS Comet and the RSMN assays, the project followed a modular validation approach (23) and was performed in three phases: Phase 0—transferability, optimisation and within-laboratory reproducibility with model genotoxins, Phase 1—between-laboratory reproducibility with three coded chemicals and Phase 2—extended the reproducibility and predictive capacity of the RSMN with a larger set of chemicals. Phase 2 for the RSMN assay had several sub-phases which was a result of interim reviews by the Steering Committee, based on the initial findings of the coded testing. Here, we investigated whether the good intra-laboratory reproducibility observed in Phase 0 and 1 with a small number of compounds (21) could be confirmed with a larger dataset, in addition to further evaluating the predictive capacity of the RSMN. Ultimately, we aimed to establish its usefulness as a 'tier 2' assay that can, alongside the RS Comet assay (1), be used to follow-up on positive results from the standard *in vitro* test battery.

Materials and methods

The RSMN assay was conducted according to the protocol previously described in detail by Dahl et al. (16). The protocol was based on the principles in the OECD (Organization for Economic Co-operation and Development) test guideline 487 for the *in vitro* MN assay (18).

Characterisation and selection of coded test chemicals

Test chemicals were selected by external experts (Dr. Raffaella Corvi and Prof. David Kirkland) from a master list prepared for Cosmetics Europe by a larger group of external experts which also included Professors Tom Slaga, Johannes Doehmer and Guenter Speit. These chemicals had previously been tested in *in vivo* genotoxicity and/or carcinogenicity studies with dermal exposure and were grouped according to three categories: true negative (TN) and true positive (TP) chemicals, with concordant *in vitro* and *in vivo* data, as well as MP for which positive *in vitro* findings were reported, but not confirmed in *in vivo* studies. Originally, a total of 38 individual chemicals were selected, comprising 17 genotoxicants (TP) with various modes of action and 21 chemicals with an expected negative outcome (TN and MP), all covering different chemical classes. During Phases 0 and 1 of the international validation of the RSMN assay (16,21),

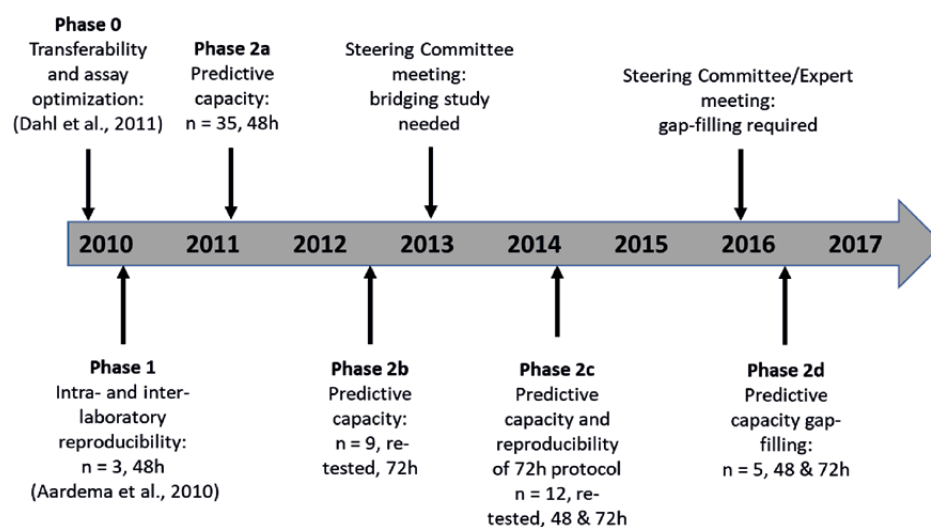


Fig. 1. RSMN assay validation timeline. Progression of the RSMN assay validation project phases over the course of the project.

a detailed protocol and harmonised scoring atlas were developed and three coded chemicals were tested in three laboratories each to demonstrate transferability and preliminarily assess intra- and inter-laboratory reproducibility (see Figure 1). A second phase focused on the predictivity of the RSMN assay and is reported here. In Phase 2a, a total of 35 chemicals were tested coded in 1–3 laboratories each using a 48-h protocol with two treatments at 24-h intervals (48-h protocol), which had previously been shown to be optimal for detecting direct DNA reactive chemicals (7,12,19,21). However, although the overall specificity (>90%) was high, the sensitivity of the 48-h protocol (65%) was lower than expected. Therefore, nine of the expected TP chemicals were retested in a follow-up study (Phase 2b) in one lab each using an extended 72-h protocol with three treatments at 24-h intervals (72-h protocol). Although the chemicals tested in this phase were not coded, the slides were randomised and coded by a separate researcher prior to scoring, so that the treatments were blinded to the scorer to prevent evaluator bias (see the section ‘Slide analysis’). The results suggested an increased sensitivity using the 72-h protocol if concentrations were carefully selected and precipitation of the test chemicals on the tissue surface was monitored, which was also separately shown for other metabolically activated genotoxins (24) as well as by an independent validation effort of a contract research laboratory (25). Following a recommendation of the external experts to more thoroughly assess the performance of the extended 72-h protocol, a Phase 2c study was conducted, in which a total of 12 coded chemicals (six TP and six MP/TN) from the previous Phases 2a and 2b were retested in two labs each. In Phase 2c, chemicals were first tested using the 48-h protocol, and any chemical that showed negative or equivocal results were retested using the 72-h protocol. Finally, an additional five coded chemicals were later chosen for testing by the external experts (Phase 2d study) to close a data gap with chemicals that were tested in the RS Comet assay (1,15), bringing the number of chemicals tested in the predictivity Phases 2a–2d of the RSMN assay validation to 40 and increasing the total number of individual coded chemicals tested to 43 overall (for Phases 1 and 2a–2d; see Figure 1).

Literature information on the *in vitro* and *in vivo* genotoxicity of the 43 chemicals included in the final evaluation are summarised in Supplementary Table S1. The test chemicals were purchased from Sigma-Aldrich or Acros Organics (Morris Plains, NJ, USA), (>95%

purity), assigned a unique code for each different phase of the validation and shipped to each laboratory by independent co-workers at either Covance Laboratories Ltd (North Yorkshire, UK), Integrated Laboratory Systems Inc. (ILS, Morrisville, NC, USA) or VitroScreen S.r.l. (Milano, Italy) at different stages of the project and each laboratory received chemicals from all subcategories (TP, TN, MP).

RSMN assay reagents

The following reagents were standardised among laboratories: CytoB, mitomycin C (MMC), 0.25% trypsin-EDTA and acridine orange (AO) were obtained from Sigma-Aldrich. Other reagents were obtained from local suppliers and were not harmonised among laboratories.

Experimental design

For each coded chemical, testing included a solubility study (assessed by visual inspection), a dose range-finding (DRF) assay (up to a maximum dose of 10% w/v, 1600 µg/cm²) and at least two valid main experiments. Test chemicals were preferably dissolved in acetone at 10% (w/v); however, if they were not soluble at this concentration, test chemicals were carefully diluted to determine their maximum solubility. If the solubility in acetone was below 1% (w/v), 70% ethanol was used instead. The results of the DRF experiment defined the maximum dose for the main experiments, depending on (a) solubility (if <10%), (b) precipitation of the chemical on the tissue surface, as seen visually at the end of the experiment or (c) excessive cytotoxicity (>60% relative to the solvent control [SC]). The concentrations tested in the main assays were targeted to induce 55 ± 5% cytotoxicity [the threshold (high) cytotoxicity limit; see below], 30 ± 10% cytotoxicity (intermediate) and 10 ± 10% cytotoxicity (low), unless the chemical showed insignificant cytotoxicity up to the maximum dose as defined by the DRF experiment. For Phases 2a and 2b, a clear positive result in the first 48-h experiment was confirmed in an independent second 48-h experiment. Depending on the outcome of the first experiment, the spacing of the test concentrations in the second experiment was modified, usually by using a tighter spacing. Beginning with Phase 2c, if the test chemical provided negative or equivocal results (see data evaluation section below for a definition of equivocal), a 72-h, three-treatment experiment was performed (Figure 2A).

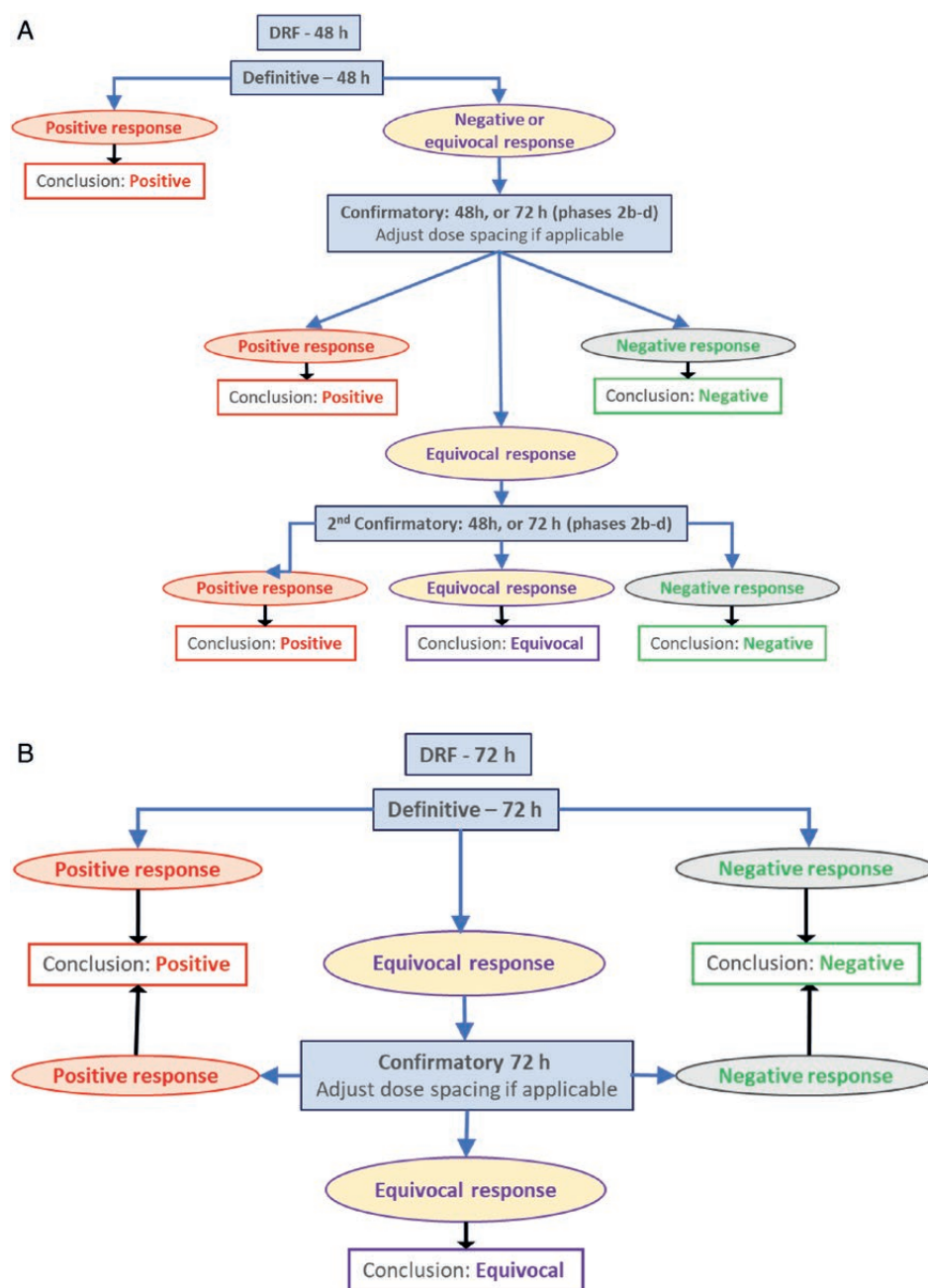


Fig. 2. (A) Decision tree for the validation exercise using exclusively 48-h protocol for Phases 1 and 2a and 48- and 72-h protocols for Phases 2b–2d. (B) Recommended decision tree for all forward-looking work, using the 72-h protocol only. Decision tree is in line with OECD 487 where clear positive or clear negative results do not need to be reproduced.

RSMN assay

The RSMN assay was conducted according to the protocol previously described in detail (16). EpiDerm™ EPI-200-MNA skin tissues and new maintenance medium (NMM) were obtained from MarTek Corporation (Ashland, MA, USA) and shipped overnight. Upon receipt, the tissues were transferred to six-well tissue culture plates containing 1 ml NMM and cultured at 37°C and 5% CO₂ for a 1-h equilibration period for the 72-h protocol or overnight for the 48-h protocol. As conducted previously, the main experiments comprised at least three valid dose groups of a test chemical in addition to a SC and a positive control (PC; MMC at 3 µg/ml in acetone).

All dose groups were tested in triplicate. Untreated negative control tissues were omitted since it had been demonstrated previously that acetone or 70% ethanol did not affect background MN levels (16). After the 1-h or overnight equilibration period, 1 ml of fresh NMM containing 3 µg/ml CytoB was added to the skin models and was refreshed every 24 h thereafter, just prior to each chemical treatment. Test chemicals were applied topically in a volume of 10 µl (equivalent to 16 µl/cm² skin surface area) to each tissue at 24-h intervals. Solutions of the test chemicals were prepared fresh daily just prior to each dosing. For the 48-h protocol, the first dose of the test chemical was applied to the tissues that had been incubated overnight and

the second dose was applied 24 h after the first dose. For the 72-h protocol, the first dose was applied 1 h after the tissues were received and the following two doses were given 24 h and 48 h after the first dose. At the end of the 48- or 72-h treatment period, keratinocytes were isolated from the skin tissues by trypsinisation and an aliquot of the cell suspension was set aside for cell counting (see below). The remaining cells were swollen in hypotonic KCl, fixed in cold methanol/acetic acid, and two slides prepared from each skin tissue as described (16).

Cytotoxicity assessment

Cytotoxicity was determined by measuring the viable cell count (VCC) and frequency of binucleated (BN) cells. The VCC from each tissue was determined by trypan blue dye exclusion or by differential staining with an AO/DAPI solution using a NucleoCounter NC250 instrument (ChemoMetec A/S, Allerød, Denmark). The BN frequency was evaluated using fluorescence microscopy of the fixed cells after AO staining (see below). The % VCC and % BN were determined by comparing the levels obtained in each treatment group with that of the SC group. The thresholds set for excessive cytotoxicity for % VCC and % BN as described previously (16) and per current OECD 487 guideline (18) for the *in vitro* MN assay were a 60% ($55 \pm 5\%$) reduction compared to the SC in either parameter. If cytotoxicity was observed with both measurements, the more sensitive parameter was used.

Slide analysis

Analysis of slides was harmonised among laboratories based on published standards (16). To ensure similar scoring of slides was achieved by each participating laboratory, a workshop was held early on in the project to discuss the scoring of micronuclei. The result of the workshop was a MN scoring atlas, which was described by Dahl et al. (16). In addition, slides from one laboratory were sent to others for re-scoring, which showed a comparable output for the slide analysis. Prior to scoring, the slides were immersed in an AO solution (40 µg/ml in DPBS) for 2–3 min, rinsed three times with DPBS and a drop of DPBS and a coverslip were placed over the slides. The fixed cells were first evaluated for % BN using a fluorescence microscope equipped with a 40× objective. Subsequently, the overall cytotoxicity of the test chemical was determined. In order to qualify for MN scoring, there must be at least two valid tissues that were not excessively cytotoxic according to the % BN or % VCC per concentration and at least three valid concentrations per test compound. If these criteria were met, then slides representing those tissues were scored for micronuclei. Prior to MN scoring, the slides were first randomised and coded by a separate person so that the treatments were blinded to the scorer to prevent evaluator or selection bias. This procedure was followed for all experiments, including the follow-up non-coded 72-h Phase 3 testing. The initial determination of % BN was measured in 500 total cells/tissue and subsequent measurement of the % of BN cells with micronuclei (% MN-BN) was determined in 1000 BN cells/tissue. For three tissue replicates, this resulted in 1500 cells per tissue analysed for % BN and 3000 cells per tissue analysed for % MN-BN. Sample size and the number of analysed cells were in line with published OECD 487 guideline recommendations (18).

Validity criteria

Prior to statistical analysis, the validity of each experiment was first determined (16). First, the experiment needed to follow the predefined experimental design: SC, PC and at least three dose groups of

the test chemical, each represented by three valid tissues with at least 500 BN cells scored per tissue. Ideally, 1000 BN cells were analysed for the presence of MN. Only results from tissues in which at least 500 BN cells were obtained were included in the statistical analysis. In addition, the yield of viable cells must be higher than 5×10^4 /skin tissue and the % BN cells in the SC must be $>25\%$. Second, the validity criteria for the control groups (SC and PC) were applied. In the main experiments, the PC must show a statistically significant increase in the % MN-BN compared to the SC. Third, a dose group was considered valid when the threshold set for cytotoxicity (i.e. a $>60\%$ reduction in either % BN or % VCC compared to the SC) was not exceeded. If excessive cytotoxicity in a dose group was observed, the dose group was not considered for the evaluation of genotoxicity. In the case of precipitation of the test chemical on top of the tissues at the end of treatment, only the lowest precipitating concentration was included in the statistical analysis as test material on the tissue surface can interfere with scoring or impact tissue quality (16).

An experiment with only two valid dose groups could nevertheless be considered valid if: (1) the first two doses were positive (in terms of genotoxicity) and the thresholds for strong cytotoxicity were not exceeded or (2) genotoxic effects were absent in all dose groups and cytotoxicity exceeded the thresholds in the third test group only (because the thresholds were set to prevent MP results). Exceptional cases, in which a single tissue may be missing (e.g. due to issues during handling), were still acceptable if a substance was clearly positive in the remaining two tissues or if the missing tissue belonged to a lower dose for a substance that did not increase % MN-BN at higher dose. Justification needed to be provided in either of these cases.

Evaluation and interpretation of results

Raw data from each single experiment were submitted in an Excel spreadsheet template for decoding and independent statistical analysis as previously described for Phase 1 (21). In addition to the raw data, the laboratories provided information on the validity of the dose groups, which could be compromised due to excessive cytotoxicity, precipitation of the test chemical at the end of the experiment on top of the skin tissues or other observations, e.g. cell abnormalities.

Once the validity of an experiment was confirmed, the MN data of the valid concentrations were analysed and interpreted according to the procedure described by Dahl et al. (16), with some clarifications. In brief, the one-sided Fisher's exact test was used to determine the statistical significance of differences between the SC and each of the test chemical treatment group for % MN-BN, where $P < 0.05$ was considered to be indicative of a statistically significant positive response, i.e. without accounting for multiple comparison. To help judge the dose dependence of the effect of experiments with equivocal results, a Cochran–Armitage (CA) trend test ($P < 0.05$) was used in the overall judgement within each laboratory. In the overall analysis after finalisation of all experimental work, the trend test from the freely available R package 'CAATexact', version 0.1.0 (available at <http://CRAN.R-project.org/package=CAATexact>) (or equivalent) was used, which we also recommend to all other users going forward. In addition, the biological relevance of the findings was considered taking into account dose dependence and strength of the effect in relation to the 95% control limit of the historical SC data as well as reproducibility between tissues within and between experiments.

A test chemical was called positive for genotoxicity overall if there were two or more concentrations of the test chemical that produced statistically significant increases in the % MN-BN outside the

upper 95% control limit of the historical SC data, as established by the 95%-quantile. The use of the 95% control limit was added to the previous RSMN evaluation criteria (16) to align with changes made to OECD guidelines (18). If only the highest concentration produced a statistically significant increase in MN induction, this effect had to be reproduced in an independent study to call a test chemical positive. If the chemical induced a significant increase at a mid-dose, it was considered positive overall if the increase was greater than the upper limit of the historical control (HC) range, the trend test was significant, and it was reproducible in an independent experiment. These criteria on effects only at a high- or mid-dose were added to clarify the criteria listed in Dahl et al. (16). A test chemical was called positive overall if it fulfilled the criteria defined for a clear positive call in either the 48-h protocol or in the extended 72-h protocol. A test chemical was called negative for genotoxicity overall if none of the above criteria for induction of MN were met in two independent studies. If the results for a test of an experiment were neither clearly positive nor clearly negative in terms of the above criteria, the biological relevance of the results was considered and, if necessary, further experimental work was conducted. For instance, if statistically significant concentrations were below the 95% control limit of the historical SC data, the experiment may be considered equivocal. Despite extensive testing, a test article may produce results that are neither clearly positive nor clearly negative. In those rare instances, the test article may be considered to have produced equivocal responses (Figure 2A).

Calculation of the predictive capacity of the RSMN assay

The predictive capacity of the RSMN assay was based on the final calls, per laboratory, for all available experimental data from all phases of the validation exercise. The predictive parameters sensitivity, specificity and accuracy were calculated from 2×2 contingency tables summarising the results per laboratory. Equivocal calls were considered as 0.5. The same parameters were calculated overall from a 2×2 contingency table obtained by summing up the calls of all laboratories accordingly.

Results and discussion

During the earlier Phase 1 of the validation, each chemical was investigated in three laboratories to obtain information on the intra- and inter-laboratory reproductivity (21). As reproducibility was found to be satisfactory, the Steering Committee instructed that each chemical could be tested in one to three laboratories for the Phases 2a–2d described here to expand the range of chemicals tested in the RSMN assay.

Results of RSMN assay validation Phases 2a–2d

Chemicals with an expected positive outcome

2-Acetylaminofluorene (2-AAF)

2-AAF was initially tested only in the 48-h protocol in Lab D in Phase 2a up to doses that caused excessive cytotoxicity. There was no significant increase in % MN-BN observed when tested up to doses inducing excessive cytotoxicity according to % VCC in two separate experiments (Supplementary Figure S1A and B; see Table 1 and Supplementary Table S2 for an overview). Similar results were also found for 2-AAF when it was retested in Phase 2c in Labs B

and C using both the 48- (Supplementary Figure S1C and E) and 72-h protocols (Supplementary Figure S1D and F). Since 2-AAF was tested up to doses producing excessive cytotoxicity without a significant response in % MN-BN, 2-AAF was incorrectly classified by all labs as negative.

2-AAF is a pro-mutagen and the known metabolic pathways involve both CYP1A2 and NAT, leading to the formation of genotoxic metabolites (26,27). In regard to genotoxicity, 2-AAF is an established mutagen and has shown positive responses in tests for both mutagenicity and clastogenicity *in vivo* in tissues with sufficient metabolic capacity for 2-AAF (see Supplementary Table S1 for an overview). In skin, there is evidence for metabolism of topically-applied 2-AAF in human skin explants (28) and a weak, but statistically significant and reproducible increase in DNA damage was detected in the RS Comet assay (1). In another study, multiple treatments of RS tissues with 2-AAF led to the induction of metabolic enzymes, generation of reactive metabolites and formation of DNA adducts, which was associated with a significant response in the RS Comet assay, but not in the RSMN assay (29). It was also demonstrated that the DNA damage caused by direct treatment with known genotoxic metabolites of 2-AAF, *N*-hydroxy-2-acetylaminofluorene and *N*-hydroxy-2-aminofluorene, was more efficiently detected in the RS Comet assay than in the RSMN assay (29).

2-Amino-3-methylimidazo[4,5-*f*]quinolone (IQ)

IQ was tested in Phase 2d in the 48-h protocol in Lab A up to doses that produced excessive cytotoxicity ($<20\%$ BN at $\geq 150 \mu\text{g}/\text{cm}^2$) in all but two doses. Since no increase in the % MN-BN was seen at either of the non-cytotoxic doses (Supplementary Figure S2A), a second experiment was conducted using the 72-h protocol. Significant increases in % MN-BN was detected in two dose groups ($\geq 57 \mu\text{g}/\text{cm}^2$) that were below the cytotoxicity threshold (Supplementary Figure S2B). Since the % MN-BN for both of these doses were outside the 95% control limit of the historical SC data, IQ was correctly classified as positive.

IQ has been shown to be an *in vivo* genotoxicant and is listed as an IARC class 2A carcinogen (i.e. reasonably anticipated to be a human carcinogen). Metabolic activation of IQ is required to cause genotoxicity and evidence suggests that IQ is sequentially *N*-hydroxylated via CYP1A2 and *O*-acetylated by NAT (30), followed by spontaneous degradation of *N*-acetoxy-IQ to an unstable DNA reactive nitrenium ion DNA (31). IQ can be transformed to a genotoxic metabolite *in vivo* in the rat (32,33); however, studies investigating the metabolism of IQ in *ex vivo* pig or human skin (over 24 h) or by EpiSkin™ S9 (over 4 h) indicate that it was not metabolised after a single treatment in either skin model (28). In this study, IQ was only found to be positive in the RSMN assay in the extended 72-h protocol, suggesting that multiple treatments are required for metabolic enzyme induction. However, IQ was not detected in the RS Comet assay after multiple treatments (15).

Azidothymidine (AZT)

AZT was initially tested in the 48-h protocol during Phase 2d in Lab A up to doses ($700 \mu\text{g}/\text{cm}^2$) that produced strong cytotoxicity. AZT caused a significant increase in the % MN-BN at multiple non-cytotoxic doses (Supplementary Figure S3A). However, since most of these increases were less than the 95% control limit of the historical SC for the lab, AZT was retested up to a higher dose ($800 \mu\text{g}/\text{cm}^2$) in the 72-h protocol. In this experiment, AZT showed significant increases in the % MN-BN that exceeded the 95% control limit of the historical SC at multiple non-cytotoxic doses (Supplementary

Table 1. Overview of validation outcome of the RSMN experiments conducted within the coded validation effort in all phases

Chemical	CAS No.	Cat	Phase	Lab A	Lab B	Lab C	Lab D	BLR
2-Acetylaminofluorene (2-AAF)	53-96-3	TP	2a,c		Neg	Neg	Neg	1
2-Amino-3-methylimidazo[4,5-f]quinolone (IQ)	76180-96-6	TP	2d	Pos				–
Azidothymidine (AZT)	30516-87-1	TP	2d	Pos				–
Cadmium chloride (CdCl ₂)	10108-64-2	TP	2a,b,c	Pos	Neg	Pos		0
Colchicine	64-86-8	TP	2a	Pos		Pos		1
Cyclopenta[<i>c,d</i>]pyrene (CPPE)	27208-37-3	TP	2a,b		Pos	Neg ^a		–
Cytosine arabinoside	147-94-4	TP	2a,b		Neg			–
2,4-Diaminotoluene (2,4-DAT)	95-80-7	TP	2a,b	Pos	Neg		Neg	0
2,3-Dibromo-1-propanol	96-13-9	TP	2a	Pos				–
Diethylstilbestrol	56-53-1	TP	2a,b		Pos			–
7,12-Dimethylbenz[<i>a</i>]thracene (DMBA)	57-97-6	TP	2d	Neg				–
Ethyl methanesulfonate (EMS)	62-50-0	TP	2a,c	Pos		Pos		1
<i>N</i> -Ethyl- <i>N</i> -nitrosourea (ENU)	759-73-9	TP	1	Pos	Pos		Pos	1
Etoposide	33419-42-0	TP	2a	Pos	Pos			1
5-Fluorouracil	51-21-8	TP	2a,b,c	Pos	Neg			0
Methyl methanesulfonate (MMS)	66-27-3	TP	2a		Pos			–
<i>N</i> -Methyl- <i>N'</i> -nitro- <i>N</i> -nitrosoguanidine (MNNG)	70-25-7	TP	2d	Pos				–
Mitomycin C	50-07-7	TP	1	Pos	Pos		Pos	1
Potassium bromate	7758-01-2	TP	2a,b,c	Pos	Pos	Pos		1
Taxol	33069-62-4	TP	2a			Pos		–
4-Vinyl-1-cyclohexene diepoxide	106-87-6	TP	2a,b,c	Pos ^b	Pos	Pos		1
Ampicillin sodium salt	69-52-3	TN	2a	Neg				–
Beclomethasone dipropionate	5534-09-8	TN	2a		Neg			–
<i>N</i> -Butyl chloride	109-69-3	TN	2a,c	Neg	Neg	Neg	Neg	1
Curcumin	458-37-7	MP	2a		Pos			–
Cyclohexanone	108-94-1	TN	1	Neg	Neg		Neg	1
2,6-Diaminotoluene (2,6-DAT)	823-40-5	MP	2a				Neg	–
2,4-Dichlorophenol	120-83-2	MP	2a	Neg			Neg	1
Diclofenac	15307-79-6	TN	2a,c	Pos	Pos		Pos	1
Ethionamide	536-33-4	MP	2a,c	Neg	Neg	Neg		1
Eugenol	97-53-0	MP	2d	Pos				–
8-Hydroxyquinoline	148-24-3	MP	2a				Neg	–
d-Limonene	5989-27-5	TN	2a,c	Neg	Neg	Neg		1
d-Mannitol	69-65-8	TN	2a		Neg		Neg	1
Nifedipine	21829-25-4	TN	2a	Neg				–
Nitrofurantoin	67-20-9	MP	2a		Neg			–
1-Nitronaphthalene	86-57-7	MP	2a	Neg				–
4-Nitrophenol	100-02-7	MP	2a,c	Neg	Neg	Neg	Neg	1
Phenanthrene	85-01-8	TN	2a,b	Pos	Neg	Neg	Neg	0
Phenol	108-95-2	MP	2a				Neg	–
Propyl gallate	121-79-9	MP	2a		Neg			–
Resorcinol	108-46-3	MP	2a,c	Equiv	Neg			0.5
Tolbutamide	64-77-7	TN	2a	Equiv	Neg	Neg		0.5

In Phase 1, each chemical was tested in three laboratories (21). Subsequently, in Phases 2a–2d, chemicals were each tested in 1–3 laboratories each to expand the chemical space tested in the RSMN assay or to expand the overlap with chemicals tested in the RS Comet assay. Cat, Category, i.e. expected outcome based on historical *in vitro* and *in vivo* genotoxicity or carcinogenicity data as provided in [Supplementary Table S1](#); MP, misleading positive; TN, true negative; TP, true positive; Pos, positive study outcome (printed in bold), DNA damage detected; Neg, negative; Equiv, equivocal; BLR, between-laboratory reproducibility.

^aCounted for predictivity but not for between-lab concordance since Lab C did not follow-up at 72 h and Lab B showed Pos at 72 h only.

^bLab A repeated the initial 48-h experiment on behalf of Lab D at 72 h, so the Lab A call was recorded as overall call.

Figure S3B). As a result of these data, AZT was correctly classified as positive.

Cadmium chloride (CdCl₂)

CdCl₂ was tested in the 48-h protocol in Lab B in Phase 2a up to doses (160 µg/cm²) that showed strong cytotoxicity in the initial DRE. No detectable increase in % MN-BN was observed in two separate experiments ([Supplementary Figure S4A and B](#)). Similarly, negative results were also observed in two subsequent experiments with the 72-h protocol in Phase 2b ([Supplementary Figure S4C and D](#)). In Phase 2c, Lab A tested CdCl₂ up to doses

that caused 50% cytotoxicity (100–160 µg/cm²) in the initial experiment with 48-h protocol without any significant increase in % MN-BN ([Supplementary Figure S4E](#)). In a subsequent experiment in Lab A using the 72-h protocol, there was a clear increase in % MN-BN ([Supplementary Figure S4F](#)). A second 72-h experiment was conducted due to the loss of tissues at several of the intermediate doses. However, the increase in the % MN-BN was not duplicated in the repeat experiment, possibly due to higher levels of cytotoxicity ([Supplementary Figure S4G](#)). Mixed results were also found in Phase 2c testing by Lab C. CdCl₂ was negative in the 48-h protocol and a follow-up 72-h experiment ([Supplementary Figure S4H and I](#)).

A repeat 72-h experiment was conducted in Lab C using closer dose spacing due to the loss of the top dose in the initial 72-h experiment as a result of excessive cytotoxicity. A significant increase in the % MN-BN was observed at the highest dose that did not exceed the cytotoxicity threshold (Supplementary Figure S4J). Overall, Lab B classified CdCl₂ as negative while Labs A and C called it positive.

The results in Lab A and C were in line with findings showing that exposure to cadmium salts is linked to an increase in mutation and MN frequencies both *in vitro* and *in vivo*, as well as to tumour formation, although conflicting results have been reported (Supplementary Table S1). Inconsistent results were observed in an *in vivo* comet assay validation study in rats exposed to daily treatment with CdCl₂ for 3 days. Initially, the results were considered positive for stomach and equivocal for liver (34); however, after re-evaluation of the data, the findings for both organs were determined to be equivocal (35). Furthermore, negative results were reported in multiple tissues in another *in vivo* comet assay study, in which mice were treated with a single i.p. injection of CdCl₂ (36). Since cadmium does not cause DNA damage in cell extracts or isolated DNA, indirect mechanisms, including the generation of reactive oxidative species and the inhibition of DNA repair have been suggested for the genotoxic effects of cadmium salts (37).

Colchicine

The maximum dose of colchicine (0.03 µg/cm²) tested was limited by excessive cytotoxicity in both Labs A and C. Colchicine was tested in Phase 2a by both labs in the 48-h protocol and highly significant increases in the % MN-BN were observed in two separate experiments each at multiple doses below the cytotoxicity threshold in Lab A (Supplementary Figure S5A and B) and Lab C (Supplementary Figure S5C and D). Due to the clear increases in the % MN-BN at doses below the cytotoxicity limit, both Labs A and C correctly classified colchicine as positive in the RSMN.

Cyclopenta[*c,d*]pyrene (CPPE)

CPPE was tested up to the maximum dose of 1600 µg/cm² in Phase 2a by Labs B and C without exceeding the cytotoxicity thresholds. In the initial 48-h experiment in Lab B, none of the doses showed a significant increase in the % MN-BN (Supplementary Figure S6A). These results were confirmed in a repeat 48-h experiment in Lab B, in which there was no significant increase in % MN-BN observed (Supplementary Figure S6B). In two separate experiments, Lab C also found negative results for CPPE using the 48-h protocol (Supplementary Figure S6D and E). CPPE was also retested by Lab B in Phase 2b using the 72-h protocol and a significant increase in % MN-BN was observed at three doses that did not exceed the cytotoxicity limits, although the response was not dose-dependent (Supplementary Figure S6C). However, due to the significant increase in % MN-BN at multiple doses, CPPE was correctly classified as positive by Lab B.

CPPE has been shown to be metabolised by several cytochrome P450 CYPs (CYP1A1, CYP1A2 and CYP3A4) to form cyclopenta[*c,d*]pyrene-3,4-epoxide, *trans*-3,4-dihydrocyclopenta[*c,d*]pyrene-3,4-diol and other DNA reactive metabolites which form DNA adducts (38). No other *in vitro* or *in vivo* MN test results for CPPE could be found. However, CPPE was positive in the Ames and mouse lymphoma assays only in the presence of metabolic activation and it has been shown to induce skin tumours in mice after dermal application (Supplementary Table S1).

Cytosine arabinoside

The maximum dose of cytosine arabinoside (48 µg/cm²) was limited by excessive cytotoxicity observed at higher doses in Lab B. Cytosine

arabinoside was initially tested in the 48-h protocol in Phase 2a and did not show a significant increase in the % MN-BN in two separate experiments (Supplementary Figure S7A and B). Negative results were seen during Phase 2b when cytosine arabinoside was retested using the 72-h protocol in Lab B (Supplementary Figure S7C) and cytosine arabinoside was incorrectly classified as negative in the RSMN assay by Lab B. Cytosine arabinoside is a nucleoside analogue that has been shown to be positive in *in vitro* MN and chromosome aberration tests in the absence of metabolic activation, as well as in *in vivo* MN and chromosome aberration tests (Supplementary Table S1).

2,4-Diaminotoluene (2,4-DAT)

The maximum dose (700 µg/cm²) of 2,4-DAT was limited by strong cytotoxicity in Lab A. In the initial experiment conducted in Phase 2a in the 48-h protocol, 2,4-DAT did not cause a significant increase in the % MN-BN; however, the % MN-BN at the two highest doses were outside the 95% control limit of the historical SC for the lab and the CA trend test was significant (Supplementary Figure S8A). After adjustment of the dose levels tested, a significant increase in the % MN-BN at multiple doses ≥420 µg/cm² that did not exceed the cytotoxicity thresholds was detected in the two additional experiments with the 48-h protocol (Supplementary Figure S8B and C). 2,4-DAT was also tested in Lab B in Phase 2a at similar dose levels, none of which caused a significant increase in % MN-BN in two separate 48-h experiments (Supplementary Figure S8D and E). Two additional experiments with 2,4-DAT in Lab B using the 72-h protocol in Phase 2b were also negative; however, the doses tested were substantially lower than those tested in the earlier 48-h experiments due to the stronger cytotoxicity observed at 72 h (Supplementary Figure S8F and G). The maximum dose of 2,4-DAT was also limited by strong cytotoxicity observed in the DRF experiment in Lab D. In two separate experiments in Phase 2a, 2,4-DAT was negative in the 48-h protocol (Supplementary Figure S8H and I). Based on the results from each lab, 2,4-DAT was classified as negative in Labs B and D and positive in Lab A.

2,3-Dibromo-1-propanol

In Lab A, 2,3-dibromo-1-propanol was tested up to a maximum dose of ~500–600 µg/cm² based on excessive cytotoxicity at higher doses. Two experiments were conducted with 2,3-dibromo-1-propanol in Phase 2a in Lab A using the 48-h protocol. In the initial experiment, a significant increase in the % MN-BN was seen only at the highest dose tested (512 µg/cm²) (Supplementary Figure S9A). However, after adjustment of the dose levels tested, a significant increase in the % MN-BN was seen at multiple doses without exceeding the cytotoxicity thresholds (Supplementary Figure S9B) and 2,3-dibromo-1-propanol was correctly classified as positive in Lab A.

Diethylstilboestrol

The maximum dose of diethylstilboestrol (16 µg/cm²) was limited by excessive cytotoxicity observed in Lab B. In two separate experiments conducted in Phase 2a, diethylstilboestrol did not significantly increase the % MN-BN using the 48-h protocol when tested at doses up to or near the cytotoxicity threshold by % VCC (Supplementary Figure S10A and B). However, when retested in the 72-h protocol in Lab B in Phase 2b, diethylstilboestrol induced a significant increase in the % MN-BN (Supplementary Figure S10C) and was correctly classified as positive in Lab B.

7,12-Dimethylbenz[*a*]anthracene (DMBA)

In Lab A, the top dose for DMBA (240 µg/cm²) was limited by precipitation on the surface of the skin tissues. DMBA did not increase % MN-BN in the first main experiment using the 48-h protocol in Phase 2d (Supplementary Figure S11A). A second experiment was then conducted using the 72-h protocol but was invalid due to excessive cytotoxicity (<40% VCC at all doses tested ≥64 µg/cm², data not shown). Based on these results, a repeat 72-h experiment was conducted after adjustment of the test doses. Excessive cytotoxicity (<40% VCC) was again observed at the highest test dose (64 µg/cm²) and DMBA did not induce an increase in % MN-BN at any of the lower valid test doses (Supplementary Figure S11B). Based on these results, DMBA was incorrectly classified as negative in Lab A.

DMBA is a pro-mutagen that requires activation by cytochrome P450 enzymes (CYP 1B1) to interact with DNA and form DNA adducts (39,40). Basal expression of phase I enzymes has been shown to be low in both native human skin and RS tissues but can be upregulated within 24–72 h (10,41). In accordance, DMBA induced DNA damage in the skin of hairless mice after a 24 h but not after a 3-h treatment period (42). DMBA has been shown to have clastogenic activity *in vitro* with metabolic activation and *in vivo* (Supplementary Table S1). DMBA was also shown to be positive in a previous RSMN study that evaluated DMBA and other pro-mutagens, although the % VCC, which was a much more sensitive indicator of the cytotoxic effects of DMBA, in the present study was not used as a measure of cytotoxicity in the earlier study and allowed higher doses of DMBA to be tested (24).

Ethyl methanesulfonate (EMS)

EMS was tested up to signs of strong cytotoxicity (360 µg/cm²) in Lab A during Phase 2a. In the initial 48-h experiment, there were no significant increases in the % MN-BN (Supplementary Figure S12A). However, in both the second and a confirmatory third 48-h experiment, EMS induced a clear and dose-dependent increase in % MN-BN at multiple doses that did not exceed the cytotoxicity limits (Supplementary Figure S12B and C). In later retesting during Phase 2c in Lab A, EMS resulted in a significant increase in the % MN-BN at one mid-dose in the initial experiment using the 48-h protocol (Supplementary Figure S12D), but the CA trend test was not significant (*P* = 0.052). EMS was then tested in the 72-h protocol, in which there was a significant increase in the % MN-BN at multiple doses below the cytotoxicity cut-off threshold (Supplementary Figure S12E). EMS was also tested up to doses inducing excessive cytotoxicity according to the % BN during Phase 2c in Lab C, in which there was a clear and dose-dependent increase in % MN-BN in the 48-h protocol (Supplementary Figure S12F). As a result of the clear increases in the % MN-BN at doses below the cytotoxicity thresholds, EMS was correctly classified as positive by both Labs A and C.

Etoposide

The top dose of etoposide was limited to 1.6 µg/cm² due to its excessive cytotoxicity at higher doses in Labs A and B. Etoposide was tested in both labs during Phase 2a and induced a clear and dose-dependent increase in % MN-BN at non-cytotoxic doses with the 48-h protocol in two separate experiments in Lab A (Supplementary Figure S13A and B) and Lab B (Supplementary Figure S13C and D). Both Lab A and Lab B correctly classified etoposide as positive.

5-Fluorouracil

In Lab A, 5-fluorouracil was tested up to the maximum dose (1600 µg/cm²) in two experiments conducted in Phase 2a in Lab

A using the 48-h protocol. There were no increases in % MN-BN in either experiment (Supplementary Figure S14A and B). Lab A retested 5-fluorouracil in Phase 2b with an adjusted dose testing range due to a much greater cytotoxicity in the 72-h DRF experiment. In two separate experiments, 5-fluorouracil induced a significant increase in % MN-BN (Supplementary Figure S14C and D). Lab A also evaluated 5-fluorouracil in Phase 2c testing and found no statistically significant increases in % MN-BN in the initial experiment with the 48-h protocol, although two doses exceeded the 95% control limit of the historical SC and the CA trend test was significant, indicating a relevant effect (Supplementary Figure S14E). In the second main experiment with the 72-h protocol, 5-fluorouracil induced a significant increase in % MN-BN at the two highest doses (Supplementary Figure S14F). As a result, Lab A correctly classified 5-fluorouracil as positive. Lab B tested 5-fluorouracil in Phase 2c and reported negative results in both the 48-h protocol (Supplementary Figure S14G) and in two separate experiments with the 72-h protocol (Supplementary Figure S14H and I) when tested up to doses that exceeded the cytotoxicity thresholds, and incorrectly classified 5-fluorouracil as negative.

Methyl methanesulfonate (MMS)

MMS was tested in Phase 2a up to the limiting cytotoxic dose (16 µg/cm²) according to % BN in Lab B. A dose-dependent increase in the % MN-BN was induced by MMS and was significant at the highest dose tested that did not produce excessive cytotoxicity with the 48-h protocol (Supplementary Figure S15A). This increase was confirmed in a repeat 48-h experiment with the same doses of MMS (Supplementary Figure S15B). Lab B correctly classified MMS as positive in the RSMN assay.

N-Methyl-N'-nitro-N-nitrosoguanidine (MNNG)

The top dose of MNNG (64 µg/cm²) was limited by excessive cytotoxicity by % BN in Lab A. MNNG was tested with the 48-h protocol and caused a clear and dose-dependent increase in % MN-BN at doses below the cytotoxicity threshold limits (Supplementary Figure S16A). Since MNNG was tested during the final Phase 2d stage and a clear positive response was observed, a confirmatory experiment was not required. MNNG was correctly by Lab A classified as positive.

Potassium bromate

Potassium bromate was tested in Phase 2a by Lab C up to the maximum test dose (1600 µg/cm²). In Lab C, in two separate experiments with the 48-h protocol, there were no significant increases in % MN-BN at doses up to 1600 µg/cm², which was near or exceeded the cytotoxicity threshold according to % VCC (Supplementary Figure S17A and B). In the Phase 2b retesting of potassium bromate in Lab C using the 72-h protocol, the cytotoxicity of potassium bromate increased significantly such that the % VCC exceeded the threshold limit at doses ≥100 µg/cm². However, after adjustment of the doses tested using the 72-h protocol, a significant increase in % MN-BN was observed at the highest dose tested (96 µg/cm²) that did not exceed the cytotoxicity threshold (Supplementary Figure S17C). Potassium bromate was also tested in Phase 2c in Labs A and B using the 48-h protocol, in which there were strong dose-dependent increases in the % MN-BN at lower doses that did not exceed the cytotoxicity threshold in Lab A (Supplementary Figure S17D) and Lab B (Supplementary Figure S17E). All three labs correctly classified potassium bromate as positive.

Taxol

Taxol was tested up to the limiting cytotoxicity dose ($1.6 \mu\text{g}/\text{cm}^2$) according to the % BN in Lab C in Phase 2a. Treatment with taxol caused a clear and dose-dependent increase in the % MN-BN at doses that did not produce excessive cytotoxicity in two separate experiments with the 48-h protocol (Supplementary Figure S18A and B). As a result of the clear and reproducible response, Lab C correctly classified taxol as positive.

4-Vinyl-1-cyclohexene diepoxide

The top dose of 4-vinyl-1-cyclohexene diepoxide was limited to $64 \mu\text{g}/\text{cm}^2$ due to excessive cytotoxicity at higher doses in Lab D. In two main experiments using the 48-h protocol during Phase 2a testing, 4-vinyl-1-cyclohexene diepoxide was negative (Supplementary Figure S19A and B). In Phase 2b, 4-vinyl-1-cyclohexene diepoxide was tested by Lab A in the 72-h protocol only. This was a follow-up to the 48-h-only protocol, tested in Lab D, at the advice of the Steering Committee, since Lab D participated in Phase 2a of the project only. There was a significant increase in % MN-BN at the highest dose tested in the first 72-h experiment (Supplementary Figure S19C). In the repeat 72-h experiment, neither dose caused a statistically significant increase in the % MN-BN, although both exceeded the 95% control limit of the historical SC and the CA trend test was also significant (Supplementary Figure S19D). During Phase 2c testing in Lab B, 4-vinyl-1-cyclohexene diepoxide induced a significant increase in % MN-BN at doses that did not exceed the cytotoxicity limits with the 48-h protocol (Supplementary Figure S19E). Lastly, in Phase 2c testing in Lab C, 4-vinyl-1-cyclohexene diepoxide was negative in an experiment using the 48-h protocol (Supplementary Figure S19F). In the initial experiment using the 72-h protocol, 4-vinyl-1-cyclohexene diepoxide showed an equivocal response with one mid-dose, resulting in a significant increase in the % MN-BN, but the CA trend test was not significant (Supplementary Figure S19G). However, in the repeat 72-h experiment, there was a significant increase in the % MN-BN at multiple doses that did not exceed the cytotoxicity limits (Supplementary Figure S19H). Labs A, which performed the 72-h follow-up testing on behalf of Lab D, as well as Labs B and C all correctly classified 4-vinyl-1-cyclohexene diepoxide as positive.

Chemicals with an expected negative outcome

This section comprises chemicals considered as 'true negative', i.e. those for which concordant negative results in historical *in vitro* and *in vivo* experiments were obtained, as well as 'misleading positive' chemicals, which showed positive *in vitro* findings, which were not confirmed *in vivo*.

Ampicillin sodium salt

Ampicillin sodium salt was tested by Lab A up to the maximum dose of $1600 \mu\text{g}/\text{cm}^2$ based on a lack of excessive cytotoxicity. The testing was carried out during Phase 2a using the 48-h protocol. There were no significant increases in the % MN-BN up to the maximum required test dose in two separate experiments (Supplementary Figure S20A and B). Ampicillin sodium salt was correctly classified as negative.

Beclomethasone dipropionate

Lab B tested beclomethasone dipropionate up to the maximum dose of $1600 \mu\text{g}/\text{cm}^2$ based on a lack of excessive cytotoxicity. The testing was carried out during Phase 2a using the 48-h protocol. There were

no significant increases in the % MN-BN up to the maximum required test dose in two separate experiments (Supplementary Figure S21A and B). Beclomethasone dipropionate was correctly classified as negative.

N-Butyl chloride

The maximum dose of N-butyl chloride tested in Phase 2a by Labs A, B and D varied from 160 to $1600 \mu\text{g}/\text{cm}^2$ according to the level of cytotoxicity observed in each lab. Each lab conducted two or three separate main experiments using the 48-h protocol and in several of the experiments, the % VCC approached or just exceeded the threshold limit at some doses. However, no significant increases in % MN-BN were observed up to the highest dose tested ($1600 \mu\text{g}/\text{cm}^2$) in any 48-h experiment [Supplementary Figure S22A and B (Lab A), S22F and G (Lab B) and S22H-J (Lab D)]. N-Butyl chloride was also re-evaluated by Labs A and C during Phase 2c testing. In the initial 48-h experiment in Lab A, there was a significant increase in % MN-BN only at the highest dose ($1600 \mu\text{g}/\text{cm}^2$) (Supplementary Figure S22C). However, two separate experiments in Lab A conducted with the 72-h protocol with the same doses showed negative results (Supplementary Figure S22D and E). The second 72-h experiment did show a significant increase in the % MN-BN at the $1600 \mu\text{g}/\text{cm}^2$ dose, but the level was within the 95% control limit of the historical SC for the lab (Supplementary Figure S22E). In Lab C, there was no significant increase in the % MN-BN in either the 48- or 72-h experiments, although the maximum dose of N-butyl chloride tested was limited to $160 \mu\text{g}/\text{cm}^2$ due to excessive cytotoxicity at higher doses (Supplementary Figure S22K and L). Due to the lack of a significant genotoxic response or the lack of a reproducible response, N-butyl chloride was correctly classified as negative in all four labs.

Curcumin

The top dose of curcumin tested ($160 \mu\text{g}/\text{cm}^2$) was limited by excessive cytotoxicity according to the % VCC in Lab B. In the first main experiment during Phase 2a, there was a marked increase in % MN-BN at three doses, although the % VCC exceeded the cytotoxicity limit at all except one dose ($48 \mu\text{g}/\text{cm}^2$) (Supplementary Figure S23A). A similar result was seen in the repeat experiment, with a significant increase in % MN-BN at $48 \mu\text{g}/\text{cm}^2$ and excessive cytotoxicity ($<40\%$ VCC) at the next dose ($96 \mu\text{g}/\text{cm}^2$; Supplementary Figure S23B). The % MN-BN only exceeded the 95% control limit of the historical SC at the lower dose, but the CA trend test was significant. On the basis of these results, curcumin was incorrectly classified as positive. Curcumin falls within the 'misleading positive' chemicals group, which has shown positive findings in the *in vitro* MN tests in most cells and cell lines but was negative when tested in an *in vivo* MN test (Supplementary Table S1).

2,6-Diaminotoluene (2,6-DAT)

In the first main experiment, 2,6-DAT was tested up to a maximum dose of $350 \mu\text{g}/\text{cm}^2$ during Phase 2a with the 48-h protocol in Lab D which was limited by excessive cytotoxicity. 2,6-DAT did not significantly increase the % MN-BN under these conditions (Supplementary Figure S24A). Due to a lack of strong cytotoxicity, the doses were adjusted up to a maximum dose of $800 \mu\text{g}/\text{cm}^2$ in the second 48-h experiment, at which 2,6-DAT caused a significant increase in % MN-BN that was outside of the 95% control limit of the historical SC at $\sim 50\%$ cytotoxicity (Supplementary Figure S24B). Since there was some inconsistency in the cytotoxicity values

amongst experiments, two additional 48-h experiments were performed. There were no significant increases in the % MN-BN in the third experiment (Supplementary Figure S24C). In the final experiment, 2,6-DAT caused a significant increase in % MN-BN at 800 µg/cm² that was outside of the 95% control limit of the historical SC. Two higher doses, up to 1000 µg/cm², did not result in a significant increase. The cytotoxicity thresholds were not exceeded and the CA trend test was also not significant (Supplementary Figure S24D). Due to the lack of reproducibility in the response of the % MN-BN, 2,6-DAT was therefore correctly predicted as negative.

2,4-Dichlorophenol

The top dose tested for 2,4-dichlorophenol (50 µg/cm²) was limited by excessive cytotoxicity according to both % BN and % VCC. In the first main 48-h experiment during Phase 2a in Lab A, no statistically significant increases in % MN-BN were observed up to this dose (Supplementary Figure S25A). This result was confirmed in a repeat experiment (Supplementary Figure S25B). In Lab D during Phase 2a testing, 2,4-dichlorophenol was negative in two experiments conducted using the 48-h protocol at similar doses (Supplementary Figure S25C and D). Therefore, 2,4-dichlorophenol was correctly classified as negative in the RSMN assay by both labs.

Diclofenac

The top dose for diclofenac (50–80 µg/cm²) was limited by excessive cytotoxicity. Diclofenac was tested in Labs B and D during Phase 2a with the 48-h protocol. In the first main experiment in Lab B, a dose-dependent increase in % MN-BN was observed, which was significant at the highest dose (50 µg/cm²) (Supplementary Figure S26A). This response, however, was not confirmed in a repeat experiment using the same doses (Supplementary Figure S26B). In the initial experiment in Lab D in Phase 2a, there was a significant increase in % MN-BN at the top dose (80 µg/cm²), however the cytotoxicity threshold according to % VCC was exceeded (Supplementary Figure S26D). A repeat 48-h experiment was conducted in which there was a significant increase in % MN-BN at the top dose of diclofenac (80 µg/cm²) that did not exceed the cytotoxicity limits (Supplementary Figure S26E). Diclofenac was also tested in Labs A and B during Phase 2c with the 48-h protocol up to similar maximum doses (80–90 µg/cm²). In these experiments, there were significant increases in the % MN-BN at all doses ≥48 µg/cm² in both labs [Supplementary Figure S26C (Lab B) and S26F (Lab A)]. Based on these results, all labs incorrectly classified diclofenac as positive.

With regards to the other available test data, diclofenac has been classified as non-genotoxic based on results of both mutagenicity [Ames, hypoxanthine (guanine) phosphoribosyl transferase (HPRT) and mouse lymphoma assay] and clastogenicity (*in vitro* chromosome aberration and *in vivo* MN) tests (Supplementary Table S1); however, only limited details of the test conditions are available. Diclofenac is a non-steroidal anti-inflammatory drug (NSAID) and also the active ingredient in dermal drugs for treating pre-cancerous skin lesions. The mechanism of action is not completely understood but likely involves the induction of apoptosis and the normalisation of metabolism and immune cell infiltration and function in actinic keratosis lesions via cyclooxygenase (COX-1/-2)-dependent and independent pathways (43–45). To note, Lab A reported a strong increase of apoptotic cells alongside the increase of micronuclei, a well-known cause of false-positive results in MN scoring. Other proposed mechanisms for diclofenac-induced toxicity include the induction of oxidative stress via metabolism of diclofenac to reactive

benzoquinone imines by CYP2C9 and CYP3A4, leading to depletion of intracellular glutathione levels and DNA damage (46,47). Along these lines, the *in vitro* chromosome aberration assay was performed in whole blood (48) and erythrocytes contain catalase which can buffer oxidative species and may explain the different outcome from that observed in the RSMN.

Ethionamide

The top dose for ethionamide (30–160 µg/cm²) varied according to the level of cytotoxicity observed in each lab. Ethionamide was tested in Lab A during Phase 2a and was shown to be negative in two separate experiments using the 48-h protocol (Supplementary Figure S27A and B). Ethionamide was also negative in Lab B when it was tested during Phase 2c up to higher doses (400 µg/cm²) that exceeded the cytotoxicity thresholds (<40% VCC or BN) using the 48- and 72-h protocols in successive experiments (Supplementary Figure S27C and D). The first experiment in Lab C with the 48-h protocol in Phase 2c was also negative when tested up to doses that exceeded the cytotoxicity cut-off limit (Supplementary Figure S27E). The follow-up experiment using the 72-h protocol produced equivocal results, with significant increases at two doses which were both at the cytotoxicity threshold limit according to the % VCC (Supplementary Figure S27F). However, in the repeat 72-h experiment conducted with the same doses of ethionamide, there was no significant increase in the % MN-BN at any dose (Supplementary Figure S27G). All three labs correctly classified ethionamide as negative in the RSMN assay.

Eugenol

Lab A tested eugenol up to doses showing excessive cytotoxicity (80–100 µg/cm²). In the first main experiment using the 48-h protocol during Phase 2d testing, eugenol induced a significant increase in % MN-BN only at one mid-dose (48 µg/cm²); however, the CA trend test was also positive (Supplementary Figure S28A). Eugenol was then retested using the 72-h protocol and a significant increase in the % MN-BN was seen at the two highest doses tested without exceeding the cytotoxicity thresholds and the response exceeded the 95% control limit of the historical SC at the highest dose (80 µg/cm²) (Supplementary Figure S28B). Based on these results, Lab A incorrectly classified eugenol as positive in the RSMN assay. Eugenol also falls within the ‘misleading positive’ chemicals group, which showed positive findings in the *in vitro* MN tests and chromosome aberration test in some cell lines but was negative when tested in *in vivo* MN tests and carcinogenicity tests (Supplementary Table S1).

8-Hydroxyquinoline

The top dose for 8-hydroxyquinoline (48 µg/cm²) was limited by excessive cytotoxicity. In the first main 48-h experiment during Phase 2a in Lab D, there were no statistically significant increases in % MN-BN up to this dose, although the cytotoxicity limits were not reached (Supplementary Figure S29A). In the second main 48-h experiment, the cytotoxicity threshold was reached according to the % BN at the top dose tested. As before, 8-hydroxyquinoline did not show any increase in % MN-BN up to this dose (48 µg/cm²) (Supplementary Figure S29B). 8-Hydroxyquinoline was classified correctly by Lab D as negative in the RSMN assay.

D-Limonene

The top dose for D-limonene (64 µg/cm²) was limited by excessive cytotoxicity observed according to the % VCC in Lab A. In two separate

48-h experiments conducted during Phase 2a, there were no statistically significant increases in % MN-BN up to this dose, although the cytotoxicity limits were not quite reached ([Supplementary Figure S30A and B](#)). D-Limonene was also tested in Phase 2c by Labs B and C using higher doses. In the initial 48-h experiment in Lab B, D-limonene did not show any increase in % MN-BN ([Supplementary Figure S30C](#)). D-Limonene was then tested using the 72-h protocol and, in the first experiment, showed a significant increase in % MN-BN at one mid-dose (680 $\mu\text{g}/\text{cm}^2$) that exceeded the 95% control limit of the historical SC and showed a significant CA trend test ([Supplementary Figure S30D](#)). However, this response was not reproducible in two separate repeat 72-h experiments at doses up to 1200 $\mu\text{g}/\text{cm}^2$ ([Supplementary Figure S30E and F](#)). Lab C tested D-limonene up to doses of 160 $\mu\text{g}/\text{cm}^2$ in the 48-h protocol and 72 $\mu\text{g}/\text{cm}^2$ using the 72-h protocol. The initial 48-h experiment showed no evidence of an increase in response to D-limonene ([Supplementary Figure S30G](#)). In the 72-h experiment, only one mid-dose (32 $\mu\text{g}/\text{cm}^2$) significantly increased the % MN-BN, exceeding the 95% control limit of the historical SC. However, the CA trend test was not significant and this response was not achieved in three higher doses of D-limonene that did not exceed the cytotoxicity limits ([Supplementary Figure S30H](#)). All three labs correctly classified D-limonene as negative.

D-Mannitol

D-Mannitol was tested up to the maximum test dose (1600 $\mu\text{g}/\text{cm}^2$) without exceeding the cytotoxicity limits in DRF experiments during Phase 2a testing in two labs. In Lab B and Lab D, D-mannitol was tested in two separate 48-h main experiments and both showed that there was no significant increase in the % MN-BN at any dose up to 1600 $\mu\text{g}/\text{cm}^2$ in Lab B ([Supplementary Figure S31A and B](#)) and in Lab D ([Supplementary Figure S31C and D](#)). D-Mannitol was classified correctly by both labs as negative.

Nifedipine

In Lab A, the top dose for nifedipine (51 $\mu\text{g}/\text{cm}^2$) was limited by precipitation on the surface of the skin tissues observed in the DRF. Nifedipine did not cause an increase in the % MN-BN in two separate 48-h main experiments conducted during Phase 2a ([Supplementary Figure S32A and B](#)). Nifedipine was classified correctly as negative.

Nitrofurantoin

In the initial 48-h experiment, nitrofurantoin was tested up to a dose of 48 $\mu\text{g}/\text{cm}^2$ due to excessive cytotoxicity at higher doses. In the first main 48-h experiment in Lab B in Phase 2a, nitrofurantoin did not cause a significant increase in the % MN-BN at any dose up to 48 $\mu\text{g}/\text{cm}^2$ ([Supplementary Figure S33A](#)). In the second main 48-h experiment, Lab B was able to retest nitrofurantoin up to the maximum test dose of 1600 $\mu\text{g}/\text{cm}^2$ without exceeding the cytotoxicity limits. None of the doses caused a significant increase in the % MN-BN ([Supplementary Figure S33B](#)). Nitrofurantoin was therefore classified correctly by Lab B as negative.

1-Nitronaphthalene

The top dose for 1-nitronaphthalene (50–80 $\mu\text{g}/\text{cm}^2$) was limited by excessive cytotoxicity observed according to the % VCC in Lab A. In the initial main 48-h experiment conducted during Phase 2a, there were no statistically significant increases in % MN-BN up to 50 $\mu\text{g}/\text{cm}^2$, although the cytotoxicity threshold limits were not reached ([Supplementary Figure S34A](#)). In the repeat 48-h experiment, 1-nitronaphthalene was tested up to a higher dose (62 $\mu\text{g}/\text{cm}^2$)

without a significant increase in % MN-BN. The response at this dose just exceeded the 95% control limit of the historical SC for Lab A and the CA trend test was significant, although the cytotoxicity threshold for % VCC was also just below the threshold ([Supplementary Figure S34B](#)). A third 48-h experiment was then conducted at doses up to 80 $\mu\text{g}/\text{cm}^2$ in which there were no significant increases in % MN-BN ([Supplementary Figure S34C](#)). As a result, 1-nitronaphthalene was classified correctly as negative in the RSMN assay.

4-Nitrophenol

4-Nitrophenol was tested up to signs of excessive cytotoxicity (20–48 $\mu\text{g}/\text{cm}^2$). Labs B and D tested 4-nitrophenol using the 48-h protocol during Phase 2a and did not find any evidence for an increase in % MN-BN in two separate experiments in each lab, although the cytotoxicity limits were not achieved at the doses tested [[Supplementary Figure S35A and B](#) (Lab B) and [S35C and D](#) (Lab D)]. 4-Nitrophenol was also tested in Phase 2c in Labs A and C using both the 48-h and follow-up 72-h protocols. All experiments were negative in both labs when tested up to doses that exceeded the cytotoxicity threshold limit by % BN [[Supplementary Figure S35E and F](#) (Lab A) and [S35 F and G](#) (Lab C)]. 4-Nitrophenol was therefore correctly classified as negative by all four labs.

Phenanthrene

The top dose for phenanthrene in the main experiments was generally limited to ~200–400 $\mu\text{g}/\text{cm}^2$ due to precipitation on the surface of the skin tissues. Phenanthrene was tested by all four labs during Phase 2a and mixed results were found. In the initial 48-h experiment in Lab A, phenanthrene induced a significant increase in % MN-BN at the two highest doses tested (160 and 224 $\mu\text{g}/\text{cm}^2$) ([Supplementary Figure S36A](#)). When tested at higher doses (up to 1600 $\mu\text{g}/\text{cm}^2$) in the second 48-h experiment, precipitation was seen at all doses ≥ 400 $\mu\text{g}/\text{cm}^2$ and the % MN-BN was not increased by any non-precipitating dose up to 270 $\mu\text{g}/\text{cm}^2$ ([Supplementary Figure S36B](#)). A negative result was also seen in the third 48-h experiment at doses up to 400 $\mu\text{g}/\text{cm}^2$ ([Supplementary Figure S36C](#)). Additional testing of phenanthrene later during Phase 2a in Lab A yielded significant increases in the % MN-BN at several non-precipitating doses in the two separate 48-h experiments ([Supplementary Figure S36D and E](#)). Phenanthrene was found to be negative in the other labs in Phase 2a in two separate 48-h experiments in each lab when tested up to precipitating doses [[Supplementary Figure S36F and G](#) (Lab B), [S36I and J](#) (Lab C) and [S36K and L](#) (Lab D)]. Finally, a 72-h experiment in Lab B during Phase 2b also showed no evidence for an increase in % MN-BN at doses of phenanthrene up to 480 $\mu\text{g}/\text{cm}^2$ ([Supplementary Figure S36H](#)). Thus, phenanthrene was incorrectly classified as positive in Lab A and correctly classified as negative in Labs B, C and D.

Phenol

Phenol was tested up to signs of excessive cytotoxicity (300–440 $\mu\text{g}/\text{cm}^2$). Lab D tested phenol using the 48-h protocol during Phase 2a, in which a significant increase in the % MN-BN was observed only at the highest dose tested (440 $\mu\text{g}/\text{cm}^2$) in the second experiment, although the % MN-BN did not exceed the 95% control limit of the historical SC for Lab D ([Supplementary Figure S37A and B](#)). On the basis of these results, phenol was correctly classified as negative in the RSMN assay in Lab D.

Propyl gallate

Propyl gallate was tested up to signs of excessive cytotoxicity (48 µg/cm²). In Lab B, using the 48-h protocol during Phase 2a, propyl gallate did not significantly increase the % MN-BN when tested up to doses that caused cytotoxicity just above the threshold limits (Supplementary Figure S38A and B). Propyl gallate was correctly classified by Lab B as negative.

Resorcinol

Lab A tested resorcinol up to signs of strong cytotoxicity (400 µg/cm²). In the first two main experiments using the 48-h protocol in Phase 2a, there were no statistically significant increases in the % MN-BN at doses up to the cytotoxicity threshold (Supplementary Figure S39A and B). Resorcinol was retested in Phase 2c by Lab A using the 48-h protocol, in which there was a significant increase in the % MN-BN at the two highest non-cytotoxic doses (Supplementary Figure S39C). Resorcinol was also tested in Phase 2c in Lab B, initially with the 48-h protocol up to doses exceeding the cytotoxicity limit by % BN. There was a significant increase in % MN-BN at the highest dose tested (160 µg/cm²) but the level did not exceed the 95% control limit of the historical SC of Lab B (Supplementary Figure S39D). In a second experiment using the 72-h protocol, resorcinol did not induce any statistically significant increase in the % MN-BN when tested up to doses that exceeded the cytotoxicity limit by % BN (Supplementary Figure S39E). Resorcinol was correctly classified as negative in Lab A in Phase 2a and Lab B in Phase 2c, but incorrectly classified as positive by Lab A in Phase 2c.

Resorcinol falls within the 'misleading positive' chemicals group and mixed results have been reported in *in vitro* MN and chromosome aberration tests in several cell types (Supplementary Table S1). Resorcinol also led to somewhat controversial results *in vivo* when it tested positive in an *in vivo* MN test in male B6C3F1 mice (49) but showed no effects in male and female Crl:CD (SD)BR rats (50) and was non-carcinogenic in rats and mice (49). Overall, both the European Food Safety Authority (EFSA) and the Scientific Committee on Consumer Safety (SCCS), the independent expert panel mandated by the European Commission, classified resorcinol as non-genotoxic *in vivo* (50,51).

Tolbutamide

The top dose for tolbutamide was limited by precipitation on the surface of the skin tissues or by excessive cytotoxicity. Tolbutamide was tested by Lab A and Lab B during Phase 2a and mixed results were found. In two separate 48-h experiments in Lab A, tolbutamide induced significant increases in the % MN-BN at two mid-doses in the initial experiment, with a significant CA trend test, and at the lowest non-precipitating dose in the second experiment. All of the significantly increased % MN-BN levels were at or just exceeded the 95% control limit of the historical SC of Lab A and were also mostly

seen in conjunction with precipitation of the chemical on the tissue surface (Supplementary Figure S40A and B). In Lab B, tolbutamide only induced a significant increase in % MN-BN at one mid-dose in the initial 48-h experiment. The CA trend test was significant but the % MN-BN at this dose did not exceed the 95% control limit of the historical SC of Lab B and was not seen in the repeat 48-h experiment (Supplementary Figure S40E and F). Tolbutamide was also retested later in Phase 2a by Lab A and Lab C and negative results were found in both labs in two separate 48-h experiments [Supplementary Figure S40C and D (Lab A) and S40 G-H (Lab C)]. Tolbutamide was incorrectly classified as positive by Lab A in early Phase 2a, but correctly classified in later Phase 2a testing and tolbutamide was also correctly classified as negative by both Labs B and C.

Historical database

For all laboratories participating in Phase 2, descriptive statistics were performed for the percentage of MN-BN for the SC (Table 2). These data were used to obtain information on the 95% quantile of the mean of the historical SC controls for the % MN-BN which was determined at the end of each test phase (2a–2d) for the 48- (two-treatment) and 72-h (three-treatment) protocols for each of the four participating laboratories. While variable among laboratories, each one (A, B, C, D) showed a low and reproducible level of the % MN-BN for the SC (as well as a clear induction of MN by the PC (MMC) across the various testing phases (2a–2d), demonstrating the robustness of the method over time. The acceptance criterion for the PC (a significant increase in % MN-BN above SC) was also fulfilled in all main experiments in Phase 2.

Assessment of intra- and inter-laboratory reproducibility

In order to assess intra- and inter-laboratory reproducibility, all data generated within the coded validation effort were tabulated from Phase 1 of the validation study published previously (21) and from Phase 2 of the present study (Table 1 and Supplementary Table S2). The reproducibility of the RSMN assay within a laboratory over time was assessed by comparing the outcome of individual experiments when more than one experiment was performed under identical conditions (using either the 48- or 72-h protocol) in the same laboratory. Across all four laboratories, 69 double and triple experiments could be identified and counted towards assessing the concordance of classification (see Table 3 and Supplementary Table S2). The overall within-laboratory reproducibility for the validation exercise was 84% (Table 3), with values between 74 and 93% for the individual laboratories that participated in the validation. Reproducibility among laboratories was calculated based on the final overall call from each lab for each chemical which was tested in multiple labs during Phases 1 and 2a–2d. In total, 22 of 43 of the

Table 2. RSMN historical solvent control for the 95%-quantile of the mean of the historical SC control for the % MN-BN determined at the end of each testing phase (2a–2d) for the 48- (two-treatment) and 72-h (three-treatment) protocols for each of the four participating laboratories (A, B, C, D; *n* = the cumulative number of experiments for the 48- and 72-h protocols)

48 h protocol	Testing phase				72 h protocol	Testing phase		
	2a	2b	2c	2d		2b	2c	2d
Lab A (<i>n</i>)	0.35 (26)	0.27 (78)	0.27 (98)	0.27 (101)	Lab A (<i>n</i>)	0.35 (25)	0.33 (33)	0.33 (37)
Lab B (<i>n</i>)	0.47 (22)	0.40 (31)	0.40 (35)	–	Lab B (<i>n</i>)	0.23 (12)	0.23 (19)	–
Lab C (<i>n</i>)	0.17 (6)	0.17 (24)	0.13 (34)	–	Lab C (<i>n</i>)	0.17 (6)	0.17 (13)	–
Lab D (<i>n</i>)	0.35 (37)	–	–	–	Lab D (<i>n</i>)	–	–	–

Table 3. Overall reproducibility within and between laboratories over time [within-laboratory reproducibility (WLR) and between-laboratory reproducibility (BLR)] in Phases 1 and 2a–2d

		Discordant	Concordant	Total	%
WLR	Lab A	6	17	23	73.9
	Lab B	3	21	24	87.5
	Lab C	1	6	7	85.7
	Lab D	1	14	15	93.3
	All labs	11	58	69	84.1
BLR		5	17	22	77.3

Table 4. Predictive capacity of the RSMN calculated based on the evaluation criteria agreed on by the Steering Committee and other external experts

Parameter	Lab A	Lab B	Lab C	Lab D	Overall
Sensitivity (%)	93.3	61.5	75.0	50.0	75.0
Specificity (%)	71.4	85.7	100	90.0	84.1
Accuracy (%)	82.8	74.1	85.7	78.6	79.8

For a per lab view, also see [Supplementary Table S1](#).

coded chemicals were tested in two or three labs each, of which 17 resulted in identical calls and 5 led to discordant final overall calls ([Tables 1 and 3](#)), leading to a between-laboratory reproducibility of 77%.

Predictive capacity of the RSMN assay

The predictive capacity of the RSMN assay was calculated using the final calls, per laboratory, for all available experimental data from all phases of the validation exercise. Applying the approach described in the Materials and methods section reveals an overall accuracy of the RSMN assay of 80%, with a sensitivity of 75% and a specificity of 84% ([Table 4](#)) (for detailed results see [Supplementary Table S3](#)).

When the RSMN assay is used in combination with the RS Comet assay as intended, which is as a follow-up to positive results from the standard *in vitro* testing battery with a choice of RSMN or RS Comet, or both, depending on the genotoxicity endpoint(s) affected ([17,52,53](#)), the sensitivity increases to 89% while the specificity remains high. A manuscript with a detailed analysis of the predictive capacity of these two assays in a ‘battery’ approach is in preparation.

Optimising the number of treatments and the overall treatment time has been an important work area for the RSMN validation project. The project initially started with a two-treatment/48-h protocol, assuming this would provide a sufficiently sensitive assay. When the initial data from Phases 1 and 2a were decoded and looked at by the steering team it was found that, although the specificity was high, the sensitivity of the 48-h protocol (65%) was lower than expected, and it was decided to retest nine of the TP chemicals in a follow-up study in one lab each using the 72-h protocol (Phase 2b). While not tested double-blind in this phase, the slides were randomised and coded prior to scoring. The results showed an increased sensitivity using the 72-h protocol compared to the 48-h protocol. To verify this and to assess specificity along with between-laboratory reproducibility, a double-blind study was conducted (Phase 2c). The ‘Phase 2c’ study was performed with 12 chemicals in two labs each and confirmed the increase in sensitivity of the 72-h protocol as there were three additional positive calls for six TPs when repeating initial negative 48-h studies (see [Supplementary Figure S1](#), experiments designated as ‘2c’). The positive impact of the 72-h protocol

on the sensitivity is in fact quite pronounced—considering the 18 calls across all laboratories for TP substances which had negative 48-h results, 11 tested positive when using the 72-h protocol (see [Supplementary Table S2](#)). This corresponds to an increase in sensitivity of 24%. Importantly, the mandatory 72-h follow-up to negative 48-h results did not lead to any false-positive calls for the six chemicals with expected negative outcome when tested in two labs each, suggesting superior overall accuracy of the 72-h protocol. An improved sensitivity of the 72-h protocol had been initially shown by Aardema ([24](#)), suggesting that such an extended protocol is especially suited for pro-mutagens needing metabolic activation. The improved sensitivity of the 72-h protocol observed in this formal validation exercise is supported by other publications. The testing of 17 chemicals of which 12 were TPs, using a different skin model (Episkin™) and a 48-h protocol, led to a sensitivity of 83% and a specificity of 60% with an overall accuracy of 76% ([54](#)). Applying a 72-h protocol to the pro-mutagens improved the sensitivity to 100% and the overall accuracy to 88%. Testing of six TPs and three non-genotoxins by a contract research lab ([25](#)) with the 48-h protocol detected all but two TPs needing metabolic activation correctly. Testing with the 72-h protocol revealed positive results also for the metabolic activation requiring chemicals B[a]P and cyclophosphamide which increased the overall accuracy to 100%. The authors suggest that greater cell division aids the improved performance of the 72-h protocol ([25](#)). Based on the consistent pattern observed in this validation and by other researchers, the RSMN validation team as well as a group of independent international experts from academia, regulatory agencies and industry ([17](#)) recommend that going forward, the experiments should start with the 72-h protocol.

Requirements for measurements of cytotoxicity

Potential cytotoxic activity of a test substance was determined by measuring the VCC and the frequency of BN cells that have completed one mitosis after the addition of the actin polymerisation inhibitor CytoB. Both are standard parameters used in the context of the *in vitro* MN test as described in test guideline OECD 487 ([18](#)). A comparison of VCC to BN to determine if there was a difference in sensitivity for measuring cytotoxicity indicated that the maximum requirement of 50–60% (OECD 487) was achieved in as many studies by VCC as by BN (see [Supplementary Figure S1](#)), demonstrating the value of both parameters to be included into the protocol. As is the case for the RS Comet assay, cytokine release can also be used to track cytotoxic properties of a test compound in the RSMN assay, as shown by Roy et al. ([55](#)).

Strategic use of the RSMN assay

The *in vitro* MN assay ([18](#)) has long become a standard assay for the detection of clastogenicity, as an alternative to the CA assay, often preferred for its ease of performance and improved ability to detect aneugenic compounds ([56](#)). This assay is typically performed using ‘2D’ standard cell cultures which have limitations in terms of how they can deal with stress, leading to a comparably high fraction of MP results ([57](#)). 3D tissue constructs, however, are thought to be more *in vivo*-like ([58–62](#)) and have also been shown to have metabolic competency similar to native human skin ([10,11,13,41](#)). This makes them ideal building blocks for this effort targeting replacement of ‘tier 2’ animal follow-up assays for the dermal exposure route which resulted in the development and validation of the RS Comet and MN assays ([1,7,14–17](#)). Both skin-based assays are considered useful tools to follow-up on positive results from the standard ‘2D’ *in vitro* genotoxicity testing battery for regulatory

decision making (9,17). Importantly, when used for this purpose, it is possible to differentiate between aneugenic and clastogenic effects by using CREST analysis, in analogy to the OECD 487 MN method, as was shown by Roy et al. (55).

Regulatory use of the RSMN was encouraged early by the SCCS, the European Commission's independent expert panel ensuring safe use of consumer products, who has added the RS genotoxicity assays to their toolbox in 2014 (63). Their most recent update of their Notes of Guidance includes a weight-of-evidence approach listing which 'tier 2' assays can be used to follow-up on positive results from the standard *in vitro* testing battery. This 'tier 2' toolbox includes the RS-based assays in the context of a dermal exposure scenario, as well as the Hen's Egg Micronucleus (HET-MN) test which is thought to better reflect the oral exposure route (64). RSMN data were already submitted and considered by the SCCS in two dossier submissions for dermally applied chemicals, a hair dye (65) and a UV filter (66).

In the meantime, several expert groups have supported the use of the 3D skin model based genotoxicity assays as 'tier 2' follow-up assays to the traditional 2D tests, including ILSI-HESI (60) and two working groups of the International Working Group on Genetic Toxicology (IWGT) at the 2009 (9) and 2017 (17) meetings. IWGT has accepted the proposed endpoint-triggered follow-up strategy and concluded that the combination of the RS Comet assay and RSMN assays resulted in highly predictive testing of genotoxicity. As a result of the favourable IWGT review which concludes that both assays are sufficiently validated to move towards the development of individual OECD test guidelines, proposals for OECD TGs for both 3D skin-based assays were accepted into OECD's test guideline development program in 2019. In this context, formal peer-review is going through ESAC, the EURL-ECVAM Scientific Advisory Committee.

Conclusions

- A total of 43 chemicals were tested in a double-blind approach to validate the RSMN assay.
- The overall within- and between-laboratory reproducibility for this validation exercise was good and showed a reproducibility of 84 and 77%, respectively, at the level of the individual experiments.
- The MatTek Epi-200™ Skin model produced a reproducibly low background MN frequency and sufficient capacity to metabolise pro-mutagens and was therefore concluded to be suitable for the RSMN assay.
- The RSMN assay was found to be predictive for *in vivo* genotoxicity testing outcomes. The overall accuracy, summarising data of four laboratories to predict a difficult set of chemicals, was 80%, with a sensitivity of 75% and a specificity of 84%. Considering endpoint specificity when following-up positives from the standard *in vitro* battery, as intended, increases the sensitivity to 89%.
- The 72-h, three-treatment protocol is superior to the 48-h protocol as shown by an increased sensitivity in detecting TP compounds.
- The RSMN is considered sufficiently validated for use as a 'tier 2' assay for dermally exposed compounds and was accepted into the OECD's test guideline development program.
- The RSMN is suggested as a follow-up tool for positive results from standard *in vitro* clastogenicity/aneugenicity assays when the expected route of exposure is dermal.

Supplementary data

Supplementary data are available at *Mutagenesis* Online.

Acknowledgments

We would like to dedicate this paper to an amazing colleague, Brenda Barnett, who recently lost her battle with cancer. Brenda was instrumental in the RSMN assay work presented here and is the person who has performed and counted more RSMN studies than anyone in the world. We would like to acknowledge not just her great work but also her wonderful attitude to life and her bravery. She will be missed.

The presented work is the result of a broad joint research project. Due to limited space, not all colleagues that contributed to this study could be added to the author list and their excellent contributions are therefore acknowledged here. These are Brigitte Faquet at l'Oréal and Nathan Wilt at IIVS. In addition, we thank Covance Laboratories (UK), Integrated Laboratory Systems, USA and VitroScreen, Italy, for their support in ordering, coding and shipping of chemicals at different stages of the project. The Cosmetics Europe Genotoxicity Task Force is thanked for their support throughout this project. Finally, the authors would like to thank Raffaella Corvi (EURL-ECVAM), David Kirkland (Kirkland Consulting), Rodger Curren and Jan van Benthem for their valuable contributions to the Steering Committee. A special thank you again to Raffaella Corvi and David Kirkland for the selection of the final validation test chemicals and their help in managing the shipment and data processing for the coded chemicals.

Funding

The work was financially supported by Cosmetics Europe—The Personal Care Association. The funding party had no influence on the study design or the generation, analysis and interpretation of data, in writing the report or in submitting the article for publication.

Conflict of interest statement: S.R., G.C.M. and R.D.C. are employed by companies that provide RSMN testing as a fee for service. All other authors declared no conflict of interest.

References

1. Pfuhler, S., Pirow, R., Downs, T. R., et al. (2020) 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* 2020 March 10 [Epub ahead of print], doi:[10.1093/mutage/geaa009](https://doi.org/10.1093/mutage/geaa009)
2. EU. (2009) European Union 2009. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products (recast). *Official J. EU*, L342, 59–209. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2009.342.01.0059.01.ENG.
3. OECD. (2015) *Test No. 430: In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test Method (TER)*, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.
4. OECD. (2015) *Test No. 435: In Vitro Membrane Barrier Test Method for Skin Corrosion*, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.
5. OECD. (2019) *Test No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method*, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.
6. OECD. (2019) *Test No. 431: In vitro skin corrosion: reconstructed human epidermis (RHE) test method*, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.
7. Curren, R. D., Mun, G. C., Gibson, D. P. and Aardema, M. J. (2006) Development of a method for assessing micronucleus induction in a 3D human skin model (EpiDerm). *Mutat. Res.*, 607, 192–204.
8. Cannon, C. L., Neal, P. J., Southee, J. A., Kubilus, J. and Klausner, M. (1994) New epidermal model for dermal irritancy testing. *Toxicol. In Vitro*, 8, 889–891.

9. Pfuhler, S., Fellows, M., van Benthem, J., *et al.* (2011) In vitro genotoxicity test approaches with better predictivity: summary of an IWGT workshop. *Mutat. Res.*, 723, 101–107.
10. Wiegand, C., Hewitt, N. J., Merk, H. F. and Reisinger, K. (2014) Dermal xenobiotic metabolism: a comparison between native human skin, four in vitro skin test systems and a liver system. *Skin Pharmacol. Physiol.*, 27, 263–275.
11. Hewitt, N. J., Edwards, R. J., Fritsche, E., *et al.* (2013) Use of human in vitro skin models for accurate and ethical risk assessment: metabolic considerations. *Toxicol. Sci.*, 133, 209–217.
12. Hu, T., Kaluzhny, Y., Mun, G. C., Barnett, B., Karetsky, V., Wilt, N., Klausner, M., Curren, R. D. and Aardema, M. J. (2009) Intralaboratory and interlaboratory evaluation of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. *Mutat. Res.*, 673, 100–108.
13. Jäckh, C., Blatz, V., Fabian, E., Guth, K., van Ravenzwaay, B., Reisinger, K. and Landsiedel, R. (2011) Characterization of enzyme activities of Cytochrome P450 enzymes, Flavin-dependent monooxygenases, N-acetyltransferases and UDP-glucuronyltransferases in human reconstructed epidermis and full-thickness skin models. *Toxicol. In Vitro*, 25, 1209–1214.
14. Reus, A. A., Reisinger, K., Downs, T. R., Carr, G. J., Zeller, A., Corvi, R., Krul, C. A. and Pfuhler, S. (2013) Comet assay in reconstructed 3D human epidermal skin models – investigation of intra- and inter-laboratory reproducibility with coded chemicals. *Mutagenesis*, 28, 709–720.
15. Reisinger, K., Blatz, V., Brinkmann, J., *et al.* (2018) Validation of the 3D Skin Comet assay using full thickness skin models: transferability and reproducibility. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, 827, 27–41.
16. Dahl, E. L., Curren, R., Barnett, B. C., *et al.* (2011) The reconstructed skin micronucleus assay (RSMN) in EpiDerm: detailed protocol and harmonized scoring atlas. *Mutat. Res.*, 720, 42–52.
17. Pfuhler, S., van Benthem, J., Curren, R., *et al.* (2020) 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). *Mutat. Res.*, 850–851, 503135.
18. OECD. (2016) *Test No. 487: In Vitro Mammalian Cell Micronucleus Test, OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris.
19. Mun, G. C., Aardema, M. J., Hu, T., Barnett, B., Kaluzhny, Y., Klausner, M., Karetsky, V., Dahl, E. L. and Curren, R. D. (2009) Further development of the EpiDerm 3D reconstructed human skin micronucleus (RSMN) assay. *Mutat. Res.*, 673, 92–99.
20. Pfuhler, S. and Reisinger, K. (2017) Part V. Skin genotoxicity. In Eskes, C., van Vliet, E. and Maibach, H. I. (eds.), *Alternatives for Dermal Toxicity Testing*. Springer International Publishing, Cham, Switzerland, pp. 507–544.
21. Aardema, M. J., Barnett, B. C., Khambatta, Z., *et al.* (2010) International prevalidation studies of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. *Mutat. Res.*, 701, 123–131.
22. Pfuhler, S., Fautz, R., Ouedraogo, G., *et al.* (2014) The Cosmetics Europe strategy for animal-free genotoxicity testing: project status up-date. *Toxicol. In Vitro*, 28, 18–23.
23. Hartung, T., Bremer, S., Casati, S., *et al.* (2004) A modular approach to the ECVAM principles on test validity. *Altern. Lab. Anim.*, 32, 467–472.
24. Aardema, M. J., Barnett, B. B., Mun, G. C., Dahl, E. L., Curren, R. D., Hewitt, N. J. and Pfuhler, S. (2013) Evaluation of chemicals requiring metabolic activation in the EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay. *Mutat. Res.*, 750, 40–49.
25. Kidd, D., Phillips, S., Chirom, T., Mason, N., Smith, R., Saul, J., Whitwell, J. and Clements, J. (2019) The 3D reconstructed skin micronucleus assay: considerations for optimal protocol design. *Mutagenesis*. [Epub ahead of print], doi:10.1093/mutage/gez037
26. Rudó, K., Meyers, W. C., Dauterman, W. and Langenbach, R. (1987) Comparison of human and rat hepatocyte metabolism and mutagenic activation of 2-acetylaminofluorene. *Cancer Res.*, 47, 5861–5867.
27. Rodrigues, A. S., Silva, I. D., Caria, M. H., Laires, A., Chaveca, T., Glatt, H. R. and Rueff, J. (1994) Genotoxicity assessment of aromatic amines and amides in genetically engineered V79 cells. *Mutat. Res.*, 341, 93–100.
28. Génies, C., Jamin, E. L., Debrauwer, L., *et al.* (2019) Comparison of the metabolism of 10 chemicals in human and pig skin explants. *J. Appl. Toxicol.*, 39, 385–397.
29. Downs, T. R., Arlt, V. M., Barnett, B. C., Posgai, R. and Pfuhler, S. (2019) Effect of 2-acetylaminofluorene and its genotoxic metabolites on DNA adduct formation and DNA damage in 3D reconstructed human skin tissue models. *Mutagenesis* 2019 December 9 [Epub ahead of print], doi:10.1093/mutage/gez044
30. Aryal, P., Terashita, T., Guengerich, F. P., Shimada, T. and Oda, Y. (2000) Use of genetically engineered *Salmonella typhimurium* OY1002/1A2 strain coexpressing human cytochrome P450 1A2 and NADPH-cytochrome P450 reductase and bacterial O-acetyltransferase in SOS/umu assay. *Environ. Mol. Mutagen.*, 36, 121–126.
31. Probst, M. R., Blum, M., Fasshauer, I., D'Orazio, D., Meyer, U. A. and Wild, D. (1992) The role of the human acetylation polymorphism in the metabolic activation of the food carcinogen 2-amino-3-methylimidazo[4,5-f]quinoline (IQ). *Carcinogenesis*, 13, 1713–1717.
32. Snyderwine, E. G., Nouse, K. and Schut, H. A. (1993) Effect of 3-methylcholanthrene induction on the distribution and DNA adduction of 2-amino-3-methylimidazo[4,5-f]quinoline (IQ) in F344 rats. *Food Chem. Toxicol.*, 31, 415–423.
33. Davis, C. D., Schut, H. A. and Snyderwine, E. G. (1993) Enzymatic phase II activation of the N-hydroxylamines of IQ, MeIQx and PhIP by various organs of monkeys and rats. *Carcinogenesis*, 14, 2091–2096.
34. Wada, K., Fukuyama, T., Nakashima, N. and Matsumoto, K. (2015) Assessment of the *in vivo* genotoxicity of cadmium chloride, chloroform, and d,l-menthol as coded test chemicals using the alkaline comet assay. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, 786–788, 114–119.
35. Uno, Y., Kojima, H., Omori, T., *et al.* (2015) JaCVAM-organized international validation study of the *in vivo* rodent alkaline comet assay for detection of genotoxic carcinogens: II. Summary of definitive validation study results. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, 786–788, 45–76.
36. Sasaki, Y. F., Sekihashi, K., Izumiyama, F., Nishidate, E., Saga, A., Ishida, K. and Tsuda, S. (2000) The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. *Crit. Rev. Toxicol.*, 30, 629–799.
37. Beyersmann, D. and Hartwig, A. (2008) Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. *Arch. Toxicol.*, 82, 493–512.
38. Kwon, H., Sahali, Y., Skipper, P. L. and Tannenbaum, S. R. (1992) Oxidation of cyclopenta[cd]pyrene by human and mouse liver microsomes and selected cytochrome P450 enzymes. *Chem. Res. Toxicol.*, 5, 760–764.
39. Buters, J. T., Sakai, S., Richter, T., *et al.* (1999) Cytochrome P450 CYP1B1 determines susceptibility to 7,12-dimethylbenz[a]anthracene-induced lymphomas. *Proc. Natl. Acad. Sci. U. S. A.*, 96, 1977–1982.
40. De Flora, S., Scarfi, S., Izzotti, A., D'Agostini, F., Chang, C. C., Bagnasco, M., De Flora, A. and Trosko, J. E. (2006) Induction by 7,12-dimethylbenz[a]anthracene of molecular and biochemical alterations in transformed human mammary epithelial stem cells, and protection by N-acetylcysteine. *Int. J. Oncol.*, 29, 521–529.
41. Hu, T., Khambatta, Z. S., Hayden, P. J., *et al.* (2010) Xenobiotic metabolism gene expression in the EpiDerm *in vitro* 3D human epidermis model compared to human skin. *Toxicol. In Vitro*, 24, 1450–1463.
42. Toyozumi, T., Ohta, R., Nakagawa, Y., Tazura, Y., Kuwagata, M., Noguchi, S. and Yamakage, K. (2011) Use of the *in vivo* skin comet assay to evaluate the DNA-damaging potential of chemicals applied to the skin. *Mutat. Res.*, 726, 175–180.
43. Nelson, C. G. (2011) Diclofenac gel in the treatment of actinic keratoses. *Ther. Clin. Risk Manag.*, 7, 207–211.
44. Singer, K., Dettmer, K., Unger, P., *et al.* (2019) Topical diclofenac reprograms metabolism and immune cell infiltration in actinic keratosis. *Front. Oncol.*, 9, 605.

45. Fecker, L. F., Stockfleth, E., Braun, F. K., Rodust, P. M., Schwarz, C., Köhler, A., Leverkus, M. and Eberle, J. (2010) Enhanced death ligand-induced apoptosis in cutaneous SCC cells by treatment with diclofenac/hyaluronic acid correlates with downregulation of c-FLIP. *J. Invest. Dermatol.*, 130, 2098–2109.
46. Owumi, S. E. and Dim, U. J. (2019) Biochemical alterations in diclofenac-treated rats: Effect of selenium on oxidative stress, inflammation, and hematological changes. *Toxicol. Res. Appl.*, 3, 2397847319874359.
47. den Braver, M. W., den Braver-Sewradj, S. P., Vermeulen, N. P. and Commandeur, J. N. (2016) Characterization of cytochrome P450 isoforms involved in sequential two-step bioactivation of diclofenac to reactive p-benzoquinone imines. *Toxicol. Lett.*, 253, 46–54.
48. EMEA. (2003) Diclofenac Summary Report EMEA/MRL/885/03-FINAL, September 2003. The European Agency for the Evaluation of Medicinal Products. Veterinary Medicines and Inspections. http://www.ema.europa.eu/docs/en_GB/document_library/Maximum_Residue_Limits_-_Report/2009/11/WC500013751.pdf. (accessed July 1, 2019).
49. NTP Testing Status of Resorcinol 10163-Y. National Toxicology Program. <https://ntp.niehs.nih.gov/testing/status/agents/rs-10163-y.html> (accessed July 1, 2019).
50. EFSA. (2010) Panel on food additives: scientific opinion on the use of resorcinol as a food additive. *EFSA J.*, 8(1):1411.
51. SCCS. (2010) Scientific Committee on Consumer Safety Opinion on Resorcinol, SCCS/1270/09. http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_015.pdf. (accessed July 1, 2019).
52. Pfuhler, S., Kirst, A., Aardema, M., et al. (2010) 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, 315–324.
53. Chen, L., Li, N., Liu, Y., et al. (2020) A new 3D model for genotoxicity assessment: EpiSkin™ Micronucleus Assay. *Mutagenesis*. [Epub ahead of print], doi:10.1093/mutage/geaa003
54. Barcham, R., Orsini, N., Andres, E., Hundt, A. and Luzy, A. P. (2018) Successful proof of concept of a micronucleus genotoxicity assay on reconstructed epidermis exhibiting intrinsic metabolic activity. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, 829–830, 75–86.
55. Roy, S., Kulkarni, R., Hewitt, N. J. and Aardema, M. J. (2016) The EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay: historical control data and proof of principle studies for mechanistic assay adaptations. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, 805, 25–37.
56. Corvi, R., Albertini, S., Hartung, T., Hoffmann, S., Maurici, D., Pfuhler, S., van Benthem, J. and Vanparys, P. (2008) ECVAM retrospective validation of in vitro micronucleus test (MNT). *Mutagenesis*, 23, 271–283.
57. Kirkland, D., Aardema, M., Henderson, L. and Müller, L. (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutat. Res.*, 584, 1–256.
58. Kubilus, J., Hayden, P. J., Ayehunie, S., Lamore, S. D. K., Servattalab, C., Bellavance, K. L., Sheasgreen, J. E. and Klausner, M. (2004) Full thickness EpiDerm™: a dermal-epidermal skin model to study epithelial-mesenchymal interactions. *Altern. Lab. Anim.*, 32, 75–82.
59. Mewes, K. R., Raus, M., Bernd, A., Zöller, N. N., Sättler, A. and Graf, R. (2007) Elastin expression in a newly developed full-thickness skin equivalent. *Skin Pharmacol. Physiol.*, 20, 85–95.
60. Zeiger, E., Gollapudi, B., Aardema, M. J., et al. (2015) Opportunities to integrate new approaches in genetic toxicology: an ILSI-HESI workshop report. *Environ. Mol. Mutagen.*, 56, 277–285.
61. Balansin Rigon, R., Kaesmeyer, S., Wolff, C., et al. (2018) Ultrastructural and molecular analysis of ribose-induced glycosylated reconstructed human skin. *Int. J. Mol. Sci.*, 19(11):3521.
62. Hausmann, C., Zoschke, C., Wolff, C., et al. (2019) Fibroblast origin shapes tissue homeostasis, epidermal differentiation, and drug uptake. *Sci. Rep.*, 9, 2913.
63. SCCS. (2014) 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. http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_156.pdf. (accessed January 7, 2020).
64. SCCS. (2018) Scientific Committee on Consumer Safety Notes of Guidance (NoG) for the Testing of Cosmetic Ingredients and Their Safety Evaluation, 10th Revision, SCCS/1602/18. https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_224.pdf. (accessed January 7, 2020).
65. SCCS. (2019) Scientific Committee on Consumer Safety Opinion on hair dye 1,2,4-trihydroxybenzene (1,2,4-THB) COLIPA no A33 (CAS 533-73-3), Submission VI, SCCS/1598/18. https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_222.pdf. (accessed January 7, 2020).
66. SCCS. (2019) Scientific Committee on Consumer Safety Opinion on Methoxypropylamino Cyclohexenylidene Ethoxyethylcyanoacetate (S87), Submission II, SCCS/1605/19. https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_227.pdf. (accessed January 7, 2020).