ESAC Opinion

on the Scientific Validity of the RS Comet and RSMN Test Methods

ESAC Opinion No. 2025-01
of 24 January 2025

Clewell, R., Corsini, E., Greco, D., Kienhuis, A., Lehmann, D., Rodríguez, B., Segner, H., Tralau, T.

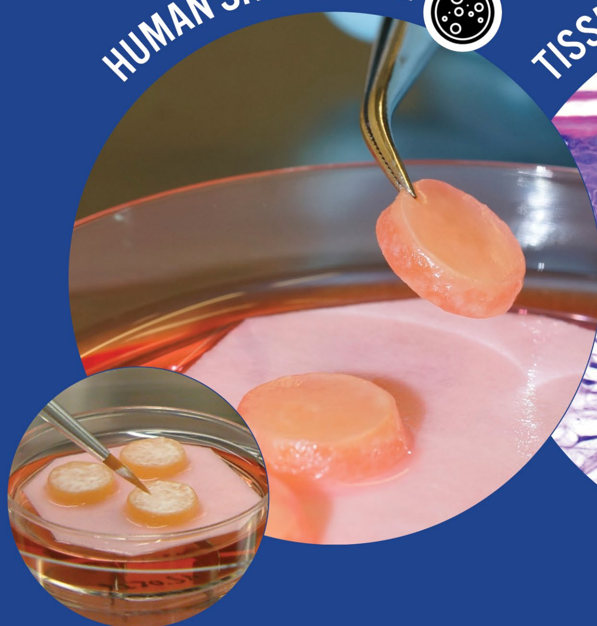
2025

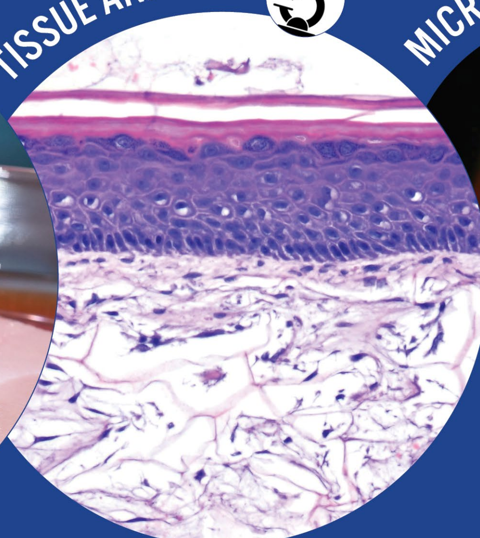
HUMAN SKIN TISSUE

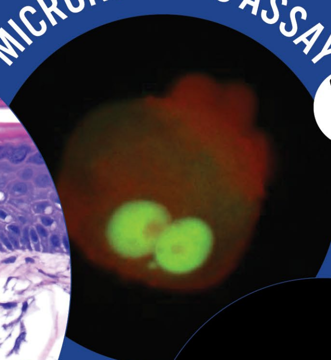
TISSUE ANATOMY

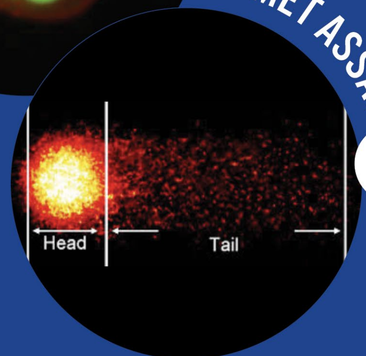
MICRONUCLEUS ASSAY

COMET ASSAY









Joint Research Centre

This document is a publication by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The independent scientific peer review of the RS Comet and RSMN *in vitro* test methods described in this report was organised by the Joint Research Centre's EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) and conducted by the EURL ECVAM Scientific Advisory Committee (ESAC).

The ESAC peer review was coordinated by João Barroso, Raffaella Corvi and Milena Mennecozzi on behalf of JRC / EURL ECVAM.

Contact information
European Commission, Joint Research Centre (JRC), Systems Toxicology Unit (F3)
Address: Via E. Fermi 2749, I-21027 Ispra (VA), Italy
Email: JRC-F3-ENQUIRIES@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC140306

PDF ISBN 978-92-68-22510-3 doi:10.2760/2520536 KJ-01-24-173-EN-N

Luxembourg: Publications Office of the European Union, 2025

© European Union, 2025

The reuse policy of the European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Unless otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of photos or other material that is not owned by the European Union permission must be sought directly from the copyright holders.

Cover page illustration, © Fira Pfuhler

How to cite this report: European Commission, Joint Research Centre, Clewell, R., Corsini, E., Greco, D., Kienhuis, A., Lehmann, D., Rodríguez, B., Segner, H. and Tralau, T., *ESAC Opinion on the Scientific Validity of the RS Comet and RSMN Test Methods*, Mennecozzi, M., Corvi, R. and Barroso, J. editors, Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/2520536, JRC140306.
Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC140306.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health and Food
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(**ESAC**)

ESAC OPINION

on the

Scientific Validity of the RS Comet and RSMN Test Methods

ESAC Opinion No.	2025-01
Relevant ESAC Request No.	2023-01
Date of Opinion	24/01/2025

Table of contents

Abstract 1
ESAC Opinion 2
Annex 1: Composition of the ESAC and ESAC Sub-Group 8
Annex 2: ESAC Sub-Group Report on RS Comet 10
Annex 3: ESAC Sub-Group Report on RSMN 40

Abstract

ESAC, the EURL ECVAM Scientific Advisory Committee, advises EURL ECVAM on scientific issues. Its main role is to conduct independent peer review of validation studies of alternative test methods and to assess their scientific validity for a given purpose. The committee reviews the appropriateness of study design and management, the quality of results obtained and the plausibility of the conclusions drawn. ESAC peer reviews are formally initiated with a EURL ECVAM Request for ESAC Advice, which provides the necessary background for the peer-review and establishes its objectives, timelines and the questions to be addressed. The peer review is normally prepared by specialised ESAC Sub-Groups. ESAC's advice to EURL ECVAM is formally provided as 'ESAC Opinions' and 'Sub-Group Reports' at the end of the peer review. ESAC may also issue Opinions on other scientific issues of relevance to the work and mission of EURL ECVAM but not directly related to a specific alternative test method.

The ESAC Opinion expressed in this report relates to the peer-review of the Reconstructed human Skin (RS) Comet and Micronucleus (RSMN) *in vitro* test methods for the assessment of genotoxicity.

Ispra, 24 January 2025

ESAC Opinion

At its 48th meeting, held on 16-17 February 2023, the EURL ECVAM Scientific Advisory Committee (ESAC) (Annex 1) was formally asked by EURL ECVAM to evaluate the scientific validity of the Reconstructed human Skin (RS) Comet and Micronucleus (MN) assays for genotoxicity testing to support their regulatory use and the development of OECD Test Guidelines. The RS Comet and RSMN are separate *in vitro* assays that use human reconstructed skin models and, because of their improved biological fidelity to human skin tissue compared to traditional submerged monoculture models, are considered better models for predicting human response to dermal chemical exposure. These assays were proposed in the Test Submissions to EURL ECVAM for use individually as follow up assays depending on the mode(s) of action represented by positive results in the traditional genotoxicity *in vitro* test battery to confirm or reject potential for genotoxicity following skin exposure.

An ESAC Sub-Group (SG) was established to assess the scientific validity of the RS Comet and RSMN assays (Annex 1). Based on its independent assessment, the ESAC SG delivered a detailed ESAC SG report for each method (Annexes 2 and 3) to support the development of this opinion. The analysis and conclusions of the ESAC SG were based primarily on a peer-review of the scientific evidence included in the RS Comet and RSMN files submitted to EURL ECVAM, including all the relevant Annexes and supporting documents (i.e., the Test Submissions). The assessment also included additional resources and supporting information provided by the Test Submitters upon request by the ESAC SG.

The final version of this opinion was unanimously endorsed by the ESAC by written procedure on 24th January 2025.

The ESAC recommends a slightly different test strategy than the one proposed by the Test Submitters. The ESAC recommends starting with the RS assay covering the mode of action that tested positive in the standard *in vitro* test battery and, if a negative result is obtained, performing the second RS assay for confirmation. A positive result in either the RS Comet or RSMN assay would be sufficient to conclude that the test item is genotoxic in the skin.

To assess the scientific validity of the RS Comet and RSMN assays, the ESAC SG evaluated:

1. RS Comet assay transferability, reproducibility, and its ability to confirm or reject positive results in traditional *in vitro* Ames or mammalian cell gene mutation tests,
2. RSMN assay transferability, reproducibility, and its ability to confirm or reject positive results in traditional *in vitro* MN tests,
3. The combined use of RS Comet and RSMN assays as a strategy to evaluate potential for dermal genotoxicity as described above.

Based on the available information, the existing scientific literature and the experts' own extensive experience as detailed in the ESAC SG reports, the ESAC unanimously concluded the following:

The RS Comet and RSMN assays are biologically relevant

The RS Comet and RSMN assays both rely on three-dimensional (3D) RS models of human origin. The RS Comet assay is performed in the Phenion® Full-Thickness Skin Model (Henkel, Düsseldorf, Germany), which is composed of mature human primary keratinocytes and fibroblasts cultured at the air-liquid-interface (ALI). The RSMN assay is performed in the EpiDerm™ Skin Model (MatTek; Ashland, MA, USA), which consists of highly differentiated normal human epithelial keratinocytes (NHEK) cultured on mesh inserts at the ALI. The assays were designed primarily for the purpose of testing dermally applied cosmetic ingredients, with the readout done at the site of contact (i.e., skin). The ESAC also sees value in potential expanded use for these tests for the evaluation of other product categories and configurations (e.g., pesticides, chemical mixtures, final products) with dermal exposure. Use of 3D tissues allows for a more biologically relevant tissue architecture, which is more likely to recapitulate *in vivo* chemical penetration and toxicological response. Both models have also been shown to express xenobiotic metabolizing enzymes, which suggests that these models have the potential to better mimic *in vivo* skin metabolism-mediated clearance and bioactivation. Thus, the ESAC considers the RS Comet and RSMN assays to offer improved biological relevance for human skin genotoxicity than traditional monoculture *in vitro* models.

RS Comet and RSMN assays offer advantages for expanding chemical domain over traditional in vitro assays

The chemical domain was not specifically explored for these assays in the two Test Submissions. However, there are clear features of a 3D skin model cultured at the ALI that promote an increased chemical domain compared to traditional genotoxicity tests performed in submerged two-dimensional (2D) monocultures. First, as previously noted, RS models have the potential to account for *in vivo* skin metabolism. Second, culturing the tissues at ALI allows the test to be used with chemicals and substances that are not generally amenable to testing in submerged *in vitro* systems, including chemicals that are

lipophilic or otherwise poorly dissolved in cell culture medium. Furthermore, complex mixtures and formulations in various vehicles (e.g., creams, emulsions) could potentially be tested in a skin model cultured at the ALI. While the Test Submissions did not specifically evaluate lipophilic chemicals or product formulations, and therefore no conclusions can be drawn on the utility of these assays for such purposes, the design of the assays does suggest potential for expanding beyond the limited chemical domain that defines all submerged 2D *in vitro* systems.

The ESAC review would have benefited from a more robust and a priori validation study design

The ESAC review of the RS Comet and RSMN assays relied on data collected and compiled over a period of more than 10 years and from multiple laboratories. This approach brings inherent challenges, some of which are clearly reflected in the Test Submissions. For example, the remit of the validation studies, as well as their design, were subject to several adjustments over time. Further, the studies included in both Test Submissions were not designed to support the rigorous Within-Laboratory Reproducibility (WLR), Between-Laboratory Reproducibility (BLR) and predictive capacity calculations that are generally performed to verify scientific validity of a proposed test method. In the RSMN assay submission, some test items were tested in only one laboratory. In other instances, chemicals were only tested in duplicate. In both the RS Comet and RSMN assays, the number of tested chemicals was low. In general, both Test Submissions would have been substantially improved by developing a validation study design a priori, and consistently following the study design. Additionally, both studies would have benefited from testing a larger number of chemicals. In particular, the ESAC recommends more testing of chemicals that are True Positives (TP) or Misleading Positives (MP) (as only these would be tested in follow-up to *in vitro* test battery results), chemicals that require bioactivation, and chemicals strategically chosen to maximize the chemical domains of the assays. To that point, the ESAC also notes that the test items were selected and classified as TP, MP and True Negative (TN) by subject matter experts that were not part of the validation study and the rationale for those decisions was not provided in the Test Submissions. Consequently, there is some uncertainty associated with these classifications that the ESAC could not resolve.

Specific test conditions or applications that did not have sufficient evidence for evaluation were identified and removed from consideration by the ESAC. Nevertheless, the RS Comet and RSMN Test Submissions were clear about their general objectives and contained the necessary controls and data, the majority of which were published and publicly available, to facilitate review by the ESAC. Further, while the design of the chemical test sets could be improved, the data provided in the Test Submission and published journal articles were sufficient for the ESAC to evaluate the scientific validity of the RS Comet and RSMN assays.

The Transferability, Reproducibility and, when the assays are used in combination as proposed by the ESAC, the Predictive Capacity of the RS Comet and RSMN are sufficient

1. RS Comet assay

Overall, the BLR of the RS Comet assay is acceptable. However, the ESAC identified concerns with respect to the transferability and the WLR. Only one test item was used to test transferability of the assay to a naïve laboratory. The ESAC considers one test item insufficient to demonstrate transferability and recommends that both TP and TN compounds are included in the evaluation of transferability. Notably, most of the chemicals used for WLR were only tested in two experiments, which does not support evaluation of WLR. Overall, the data set suffered from having a small number of tests. The predictive capacity varied widely across laboratories, with one laboratory testing too few chemicals to accurately evaluate the predictive capacity. The laboratory that tested the most chemicals (n = 18) demonstrated sensitivity and specificity greater than 80%. The ESAC performed analyses in which laboratories with insufficient tests were excluded and the BLR data was used to bolster the WLR and transferability data. Based on these analyses, and the submitted BLR, the ESAC concludes that the results of the validation study regarding reproducibility, transferability and predictive capacity are adequate for the defined purpose.

The RS Comet assay was able to identify four out of seven TP chemicals that require bioactivation for genotoxicity, demonstrating that this assay may exhibit metabolic competence needed to identify genotoxicants that require bioactivation. However, only one bioactive TP test item was tested in more than one laboratory, limiting the ESAC's ability to evaluate consistency between laboratories. Therefore, additional data should be collected to support this application of the assay.

2. RSMN assay

Overall, the predictive capacity, WLR and BLR of the chemicals tested at 48 h is adequate, with exception of the test items that require metabolic activation. Of the seven TP chemicals tested to evaluate the ability of the RSMN assay to identify bioactivated genotoxicants, only three were tested in more than one laboratory and only one chemical showed a positive response in all three test laboratories. Based on the data provided, the ESAC concluded that the metabolic competence of the RSMN assay may not be sufficient to identify bioactivated genotoxicants, possibly due to immature keratinocytes. While a number of peer-reviewed publications were provided with the Test Submission to support the expression of native metabolizing enzymes in the EpiDerm tissue, based on the current validation data provided, the ESAC cannot draw conclusions on the utility of the RSMN alone for testing potentially bioactivated genotoxicants.

While the Test Submission recommended using the 72-h time point for the RSMN assay rather than the 48-h time point initially tested by the validation laboratories, the ESAC found that there were insufficient data to evaluate the 72-h time point based on the Test

Submission. However, in a recent publication (Thakkar et al., 2022)[1], the added value of the 72-h time point was demonstrated with a set of 22 fragrance materials. Based on the combined data, the ESAC concurs with the use of the 72-h time point for the RSMN assay and considers the strategy proposed appropriate.

3. Combined RS Comet and RSMN

While the current data was insufficient to comprehensively evaluate the utility of either the RS Comet or RSMN assay alone to identify bioactivated genotoxicants, the ESAC found that when results of the RS Comet and RSMN assays were combined, all but one of the seven bioactive TP test items tested in both RS Comet and RSMN were correctly classified. The only outlier (i.e., diaminotoluene) was negative in the RS Comet and positive in one out of three laboratories in the RSMN. Therefore, the ESAC concludes that the combination of RSMN and RS Comet is preferred for general use and particularly for identifying substances that require bioactivation.

ESAC comments on the utility of the RS Comet and RSMN assays in regulatory testing

According to the Test Submitters, the RS Comet and RSMN assays were designed to follow up the results of a standard *in vitro* genotoxicity test battery, in the context of test items that are primarily associated with dermal exposure. The Test Submitters also stated that the RS assays should not be used to assess genotoxicity of compounds via other routes of exposure (e.g. oral). The ESAC sees value in potential expanded use for these tests for the evaluation of other product categories and configurations (e.g., pesticides, chemical mixtures, final products). The ESAC also considers that these tests could add value to a WoE-based evaluation of systemic genotoxicity provided systemic bioavailability is considered. It should be noted that, in the opinion of the ESAC, the RS assays will not allow any predictions on potential systemic genotoxicity. Further, the ESAC notes that accounting for systemic genotoxicity will require additional considerations such as the use of liver S9 fraction and/or data on skin absorption/penetration. In cases of significant skin absorption of a compound found to be positive in traditional genotoxicity assays following metabolic activation (by S9) but negative in the RS assays, the possibility of its liver activation into a genotoxic compound cannot be excluded due to potential differences in the metabolic capacity of the skin and the liver. The use of reconstructed 3D skin models not only offers the advantage of organ similarity with regard to the barrier function but also (limited) metabolic competence. Further, unlike other *in vivo* animal and non-human cell-based test systems, both RS assays also offer the advantage of human relevance. The ESAC recognizes that current approaches to validation of test methods do not specifically acknowledge human relevance and is of the opinion that this is an important point that

---

[1] Thakkar Y, Moustakas H, Aardema M, Roy S, Pfuhler S, Api AM (2022) Use of the EpiDerm™ 3D reconstructed skin micronucleus assay for fragrance materials. Mutagenesis 37(2):89-111. doi: 10.1093/mutage/geab040

should be given more weight in future validations. Hazard classification should be straightforward for all substances identified as initial positives which are not subject to S9-activation as these will either be confirmed or overruled by subsequent testing with RS assays. However, for substances originally requiring S9-activation, overruling by the RS assays is less straightforward as a negative result obtained using the RS assays cannot principally exclude the possibility of systemic activation following skin penetration.

Summary and Conclusions

The ESAC concludes that the evidence provided on the RS Comet assay is sufficient and adequate to support its scientific validity for use with chemicals that test positive in the Ames or mammalian cell gene mutation tests, including chemicals that require bioactivation.

The ESAC further concludes that the evidence provided on the RSMN assay 72-h time point is sufficient and adequate to support its scientific validity for use with chemicals that are directly genotoxic and test positive in the standard *in vitro* micronucleus test (submerged monoculture). However, the ESAC cautions that insufficient data exist to evaluate the scientific validity of the test for use with bioactivated genotoxic chemicals. Additionally, while the Test Submission recommended using the 72-h time point for the RSMN assay rather than the 48-h time point initially tested by the validation laboratories, the ESAC found that there were insufficient data to evaluate the 72-h time point based on the Test Submission. However, in a recent publication (Thakkar et al., 2022)[1], the added value of the 72-h time point was demonstrated with a set of 22 fragrance materials. Based on the combined data, the ESAC concurs with the use of the 72-h time point for the RSMN assay, and considers the strategy proposed appropriate.

The ESAC concludes that these are valuable assays as part of a tiered testing strategy, including using these RS assays as follow-up tests to the standard *in vitro* test battery for genotoxicity in skin following dermal application. For chemicals in which dermal genotoxicity is the toxicological endpoint of concern following application to the skin, the ESAC recommends starting with the RS assay covering the mode of action that tested positive in the standard *in vitro* test battery. In case of a negative result, the second RS assay should be performed for confirmation. A positive result in either the RS Comet or RSMN assay would be sufficient to conclude that the test item is genotoxic in the skin. The ESAC also notes the promise of these tests to contribute to an animal-free New Approach Methodology (NAM)-based approach to systemic genotoxicity following dermal exposure, with additional development and as part of a WoE approach.

Annex 1

COMPOSITION OF THE ESAC AND ESAC SUB-GROUP

Composition of the ESAC and ESAC Sub-Group

EURL ECVAM Scientific Advisory Committee (ESAC)

- Dr. Rebecca CLEWELL (ESAC Chair)
- Prof. Emanuela CORSINI
- Prof. Dario GRECO
- Dr. Sebastian HOFFMANN (excluded from review due to a conflict of interests)
- Dr. Anne KIENHUIS
- Dr. David LEHMANN
- Prof. Blanca RODRÍGUEZ
- Prof. Helmut SEGNER
- Dr. Tewes TRALAU

ESAC Sub-Group

- Prof. Rebecca CLEWELL (Sub-Group Chair)
- Prof. Emanuela CORSINI
- Dr. Anne KIENHUIS
- Dr. David LEHMANN
- Prof. Blanca RODRÍGUEZ
- Prof. Helmut SEGNER
- Dr. Tewes TRALAU

EURL ECVAM (Secretariat)

- Dr. João BARROSO (ESAC Coordinator)
- Dr. Raffaella CORVI
- Dr. Milena MENNECOZZI
- Prof. Maurice WHELAN (Head of Unit)

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health and Food
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Annex 2

ESAC SUB-GROUP REPORT ON RS COMET

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health and Food
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(**ESAC**)

ESAC SUB-GROUP REPORT

on the

Scientific Validity of the RS Comet Test Method

Title page information			
File name	ESAC_SG_Report_RSCOMET.docx		
Abbreviated title of ESAC request	RSCOMET-RSMN		
Relating to ESAC REQUEST Nr.	2023-01		
Request discussed through	ESAC 48 (February 2023)		
Report to be handed over to ESAC Chair and EURL ECVAM Coordinator by	Rebecca Clewell (Sub-Group Chair)		
Version tracking			
Date	**Version**	**Author(s)**	**Description**
18/12/2023	V1.0	ESAC SG	First agreed draft of ESAC SG Report
23/05/2024	V2.0	ESAC SG	Final approved draft of ESAC SG Report sent to ESAC for endorsement
24/01/2025	V3.0	ESAC	Final version of ESAC SG Report endorsed by the ESAC

Table of Contents

TABLE OF CONTENTS 12
ESAC SUB-GROUP (SG) 14
ABBREVIATIONS USED IN THE DOCUMENT 15
1. STUDY OBJECTIVE AND DESIGN 16
1.1 Analysis of the clarity of the study objective's definition 16
(a) ESAC SG summary of the study objective as outlined in the Test Submission 16
(b) Appraisal of clarity of study objective as outlined in the Test Submission 16
1.2 Quality of the background provided concerning the purpose of the test method 17
(a) Analysis of the scientific rationale provided in the Test Submission 17
(b) Analysis of the regulatory rationale provided in the Test Submission 18
1.3 Appraisal of the appropriateness of the study design 19
1.4 Appropriateness of the statistical evaluation 19
2. COLLECTION OF EXISTING DATA 20
2.1 Existing data used as reference data 20
2.2 Existing data used as testing data 20
2.3 Search strategy for retrieving existing data 20
2.4 Selection criteria applied to existing data 20
3. QUALITY ASPECTS RELATING TO DATA GENERATED DURING THE STUDY 21
3.1 Quality assurance systems used when generating the data 21
3.2 Quality check of the generated data prior to analysis 21
4. QUALITY OF DATA USED FOR THE PURPOSE OF THE STUDY (EXISTING AND NEWLY GENERATED) 22
4.1 Overall quality of the evaluated testing data (newly generated or existing) 22
4.2 Quality of the reference data for evaluating relevance 22
4.3 Sufficiency of the evaluated data in view of the study objective 22
5. TEST DEFINITION (MODULE 1) 23
5.1 Quality and completeness of the overall test definition 23
5.2 Quality and completeness of the documentation concerning protocols and prediction models 23
6. TEST MATERIALS 24
6.1 Sufficiency of the number of evaluated test items in view of the study objective 24
6.2 Representativeness of the test items with respect to applicability 24
7. WITHIN-LABORATORY REPRODUCIBILITY (WLR) (MODULE 2) 25
7.1 Assessment of repeatability and reproducibility in the same laboratory 25
7.2 Conclusion on within-laboratory reproducibility as assessed by the study 25
8. TRANSFERABILITY (MODULE 3) 26
8.1 Quality of design and analysis of the transfer phase 26
8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study 26
9. BETWEEN-LABORATORY REPRODUCIBILITY (BLR) (MODULE 4) 27
9.1 Assessment of reproducibility in different laboratories 27
9.2 Conclusion on between-laboratory reproducibility as assessed by the study 27
10. PREDICTIVE CAPACITY AND OVERALL RELEVANCE (MODULE 5) 28
10.1 Adequacy of the assessment of the predictive capacity in view of the purpose 28
10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose 28

11. APPLICABILITY DOMAIN (MODULE 6) 29
11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions 29
11.2 Quality of the description of applicability domain, limitations, exclusions 30
12. PERFORMANCE STANDARDS (MODULE 7) 31
12.1 Adequacy of the proposed Essential Test Method Components 31
12.2 Adequacy of the proposed Reference Chemicals 31
12.3 Adequacy of the proposed performance target values 31
13. READINESS FOR STANDARDISED USE 32
13.1 Assessment of the readiness for regulatory purposes 32
13.2 Assessment of the readiness for other uses 32
13.3 Critical aspects impacting on standardised use 33
13.4 Gap analysis 33
14. OTHER CONSIDERATIONS 34
15. CONCLUSIONS ON THE STUDY 35
15.1 ESAC SG summary of the results and conclusions of the study 35
15.2 Extent to which study conclusions are justified by the study results alone 36
15.3 Extent to which conclusions are plausible in the context of existing information 36
16. RECOMMENDATIONS 37
16.1 General recommendations 37
16.2 Specific recommendations (e.g., concerning improvement of protocols) 37
17. REFERENCES 38

ESAC Sub-Group (SG)

Full title: ESAC Sub-Group on RS Comet and RSMN Assays

Abbreviated title: ESAC SG RSCOMET-RSMN

The European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) Scientific Advisory Committee (ESAC) Sub-Group (SG) was established in February 2023 during the 48th plenary meeting of the ESAC (ESAC48) to assist in the production of an ESAC Opinion on the scientific validity of the Reconstructed human Skin (RS) Comet and Micronucleus (MN) *in vitro* test methods for the assessment of genotoxicity.

This ESAC SG Report on the scientific validity of the RS Comet test method was prepared at the request of EURL ECVAM by the "ESAC Sub-Group on RS Comet and RSMN Assays" (ESAC SG), which was charged with conducting a detailed scientific peer review of the external validation studies of the RS Comet and RSMN *in vitro* test methods. The basis for the scientific peer review was the EURL ECVAM Request for ESAC Advice approved by the ESAC during the ESAC48 plenary meeting of February 2023 (ESAC request 2023-01).

The ESAC SG met at the Joint Research Centre (JRC) in Ispra on 22-23/06/2023 and virtually on 13 and 26/10/2023; 8, 16, 20 and 24/11/2023; 1, 4, 11 and 18/12/2023; 10 and 22/01/2024; 5 and 14/02/2024; 1 and 22/03/2024; 3 and 6/05/2024, to conduct its peer review. The ESAC SG Report on RS Comet was endorsed by the ESAC SG on 23/05/2024 and represents its consensus view. The Report was slightly revised and finally endorsed by the ESAC on 24/01/2025 following meetings on 8 and 20/01/2025.

The ESAC SG had the following members:

- Dr. Rebecca CLEWELL (SG Chair)
- Prof. Emanuela CORSINI
- Dr. Anne KIENHUIS
- Dr. David M. LEHMANN
- Prof. Blanca RODRÍGUEZ
- Prof. Helmut SEGNER
- Dr. Tewes TRALAU

EURL ECVAM (Secretariat):

- Dr. João BARROSO (ESAC Coordinator)
- Dr. Raffaella CORVI
- Dr. Milena MENNECOZZI

ABBREVIATIONS USED IN THE DOCUMENT

- **2D** Two-dimensional
- **3D** Three-dimensional
- **ANOVA** Analysis of variance
- **BLR** Between-laboratory reproducibility
- **CA** Chromosomal aberration test
- **DSB** Double-strand break
- **ESAC** EURL ECVAM Scientific Advisory Committee
- **ESAC SG** ESAC Sub-Group
- **EURL ECVAM** European Union Reference Laboratory for Alternatives to Animal Testing
- **JRC** Joint Research Centre
- **MN** Micronucleus
- **MNT** Micronucleus Test
- **MP** Misleading Positive
- **OECD** Organisation for Economic Co-operation and Development
- **PM** Prediction Model
- **REACH** Registration, Evaluation, Authorisation and Restriction of Chemicals
- **RS** Reconstructed human Skin
- **SCCS** Scientific Committee on Consumer Safety
- **SG** Sub-Group
- **SOP** Standard Operating Procedure
- **SSB** Single-strand break
- **TG** Test Guideline
- **TN** True Negative
- **TP** True Positive
- **TST** Test Submission Template (the Full Test Submission)
- **Vs** Versus
- **WLR** Within-laboratory reproducibility
- **WoE** Weight-of-Evidence

1. Study objective and design

1.1 Analysis of the clarity of the study objective's definition

(a) ESAC SG summary of the study objective as outlined in the Test Submission

The Reconstructed human Skin Comet (RS Comet) assay is an animal-free adaptation of the alkaline Comet Assay (OECD TG 489 (OECD, 2016a)) that addresses the potential of a substance to cause genotoxicity in the form of DNA strand breaks, which can lead to clastogenic effects, and DNA lesions (apurinic/apyrimidinic site) that can lead to gene mutations. The RS Comet assay has been adapted to a three-dimensional (3D) Reconstructed human Skin (RS) model, namely the Phenion® Full-Thickness Skin Model (Henkel, Dusseldorf, Germany). The model is composed of human primary keratinocytes and fibroblasts cultured under air-liquid-interface (ALI) conditions on a bovine collagen scaffold. The RS Comet assay was initially developed using EpiDerm™ EPI-200-MNA, the same used for the Reconstructed human Skin Micronucleus (RSMN; see RSMN Sub-Group report) but was ultimately transitioned to the Phenion® Full-Thickness Skin Model, as the use of full-thickness RS resulted in fewer non-qualified experiments compared to the 3D human epidermal skin model.

Genotoxicity testing usually follows a tiered approach, with higher tier testing routinely being conducted *in vivo* as a follow up of *in vitro* positive results. This approach creates challenges for substances where *in vivo* testing is banned, such as cosmetic ingredients. It also limits testing efficiency, negatively affects animal welfare, and fails to address societal expectations for more humane testing.

The RS Comet Test Submission Template (TST) presents the assay as part of a tiered testing strategy for the evaluation of genotoxicity following dermal exposure. The respective strategy proposes the use of the RS Comet assay as a follow up to confirm or reject a positive Ames and/or mammalian cell gene mutation tests, and the RSMN assay (see Annex 3: RSMN Sub-Group Report) as a follow up to confirm or reject a positive *in vitro* micronucleus test (MNT) and/or chromosomal aberration test (CA). While this proposed testing battery will not inform the underlying mechanisms of genotoxicity or possible substance or metabolite absorption, it nevertheless covers all possible types of DNA damage (i.e., gene mutations, clastogenicity, and aneugenicity). The RS Comet and the RSMN assays both rely on 3D RS models with the proposed strategy aimed at regulatory use in the context of cosmetic ingredient testing for genotoxicity associated with skin exposure. However, the tests also have potential utility for the testing of pesticides, substances under Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), or pharmaceutical products, applied topically. Herein, the ESAC provides a review of both the RS Comet and RSMN TSTs.

For chemicals in which dermal genotoxicity is the toxicological endpoint of concern following application to the skin, the ESAC recommends using both assays as a follow up to any positive result in the standard genotoxicity *in vitro* test battery. A positive result in either the RS Comet or RSMN would be sufficient to conclude that the test item is genotoxic. The ESAC recommends starting with the RS assay covering the mode of action that tested positive in the standard *in vitro* test battery and only performing the second RS assay in case of a negative result in the first RS assay.

(b) Appraisal of clarity of study objective as outlined in the Test Submission

The review of the RS Comet assay by the ESAC relies on data collected and compiled over a period of more than 10 years. This approach brings inherent challenges, some of which are clearly reflected in the TST. For example, the remit of the validation study, as well as its design, were subject to several adjustments over time (see below).

The Test Submitters were able to retrospectively address most of these issues, although some uncertainties remain about the reasoning behind some of the test substance selections as well as the justification for some of the protocol adjustments. Regarding the chemical space covered in the validation studies, the TST would have benefitted from additional details related to the underlying

mechanisms, substance class representation, and stability of substances selected for the validation study. For example, only a small number of substances included in the validation study require bioactivation (i.e., 7 of which 3 were also evaluated during assay development and optimization). Furthermore, the strategy for chemicals selection has not been clearly described or justified relative to the context of use described in the TST or in the supporting documents. To this point, there appears to be a disconnect between the proposed purpose of the assay (i.e., confirmation of positive results obtained in the standard *in vitro* genotoxicity battery) and the chemicals selection strategy. A balanced set of true positive (TP), true negative (TN), and misleading positive (MP) substances would seem to be more appropriate for a standalone use of the methods rather than the proposed use of confirming positive results. These limitations are addressed in later sections of this report (see subsections 6.2 and 11.2).

Nevertheless, the TST is clear about its general objective and contains the necessary controls and data, the majority of which is published and publicly available, to facilitate review by the ESAC. Further, while a more context-specific test item set would be preferred, the substances tested during the validation study included TP, TN, and MP. Likewise, metabolic competency was tested using a selection of chemicals that require bioactivation. Thus, the TST contains sufficient information for evaluation of the RS Comet assay and clearly outlines the potential regulatory benefit of its application.

1.2 Quality of the background provided concerning the purpose of the test method

The TST adequately describes the purpose of the assay (i.e., follow-up test for substances testing positive in the Ames and/or mammalian cell gene mutation tests, and for which genotoxicity was the toxicological endpoint of concern following topical application to the skin). The method design concept involved merging conventional genotoxicity parameters with existing 3D, RS technology. This integration led to the development of the "3D" RS Comet assay, discussed in more detail here, and an RSMN assay (see Annex 3: RSMN Sub-Group Report). The combined use of these assays is considered suitable for investigating positive results obtained from standard two-dimensional (2D) *in vitro* genotoxicity assays (i.e., Ames test, mammalian cell gene mutation test, MNT and CA). Depending on the outcomes of the standard *in vitro* battery, which typically covers all genotoxicity endpoints (i.e., gene mutation, clastogenicity, and aneugenicity), a choice would be made between the RS Comet assay or RSMN assay (or both).

In recent decades, the use of animal testing for chemicals has come under increased scrutiny, prompting a global shift away from such studies due to concerns for animal welfare. Reflecting this, the European Union implemented the 7th Amendment to the Cosmetics Directive of the European Commission in 2003, prohibiting the use of *in vivo* follow-up testing for cosmetic ingredients since March 2009. In the context of a 'test battery,' this implies that a positive outcome from an *in vitro* standard genotoxicity assay would render the ingredient unsuitable for use in cosmetic products. Contemporary OECD genotoxicity Test Guidelines (TGs), such as OECD TGs 474 (OECD, 2016b), 488 (OECD, 2022), and 489 (OECD, 2016a), now underscore the importance of considering the intended or expected route of human exposure. Additionally, there is a growing emphasis on testing the 'site-of-contact,' as outlined in OECD TGs 488 (OECD, 2022) and 489 (OECD, 2016a). This is particularly relevant for ingredients in cosmetics, household products, pesticides, where the primary site of exposure is often the skin. All these points support the development and application of the RS Comet and RSMN assays.

(a) Analysis of the scientific rationale provided in the Test Submission

While the RS Comet assay TST provides a cursory overview of the scientific rationale for test method development, the submission relies heavily on information previously published in scientific journals by the Test Submitters and the corresponding standard operating procedure (SOP) for the method. This format for a TST poses substantial challenges for reviewers because many information sources

must be reviewed to properly assess the test method under consideration. Future submissions would benefit from including a summary of key information on scientific rationale, regulatory need, and assay performance in the main TST document. Nonetheless, the submission package, which includes the TST and several informative publications, provide the necessary scientific rationale for the validation of this assay.

The neutral Comet assay is a single cell gel electrophoresis assay used to detect DNA double strand breaks (DSBs). The alkaline version of the Comet assay can detect both single strand breaks (SSBs) and DSBs. Thus, the alkaline Comet assay is a more comprehensive measure of chemical-induced DNA damage. Strand breaks may result from direct interaction of the test chemical with DNA, as well as incomplete DNA repair or can occur at alkali labile sites. In its alkaline version, the Comet assay covers DNA damage which may lead to clastogenic damage and potentially gene mutation lesions. Notably, and although mutagenicity might be the result of a variety of mechanisms, the alkaline Comet Assay is reported to have a high predictive capacity for this endpoint (Kirkland et al., 2008).

Apart from its ease of use, all versions of the Comet assay have the inherent technical advantage of straightforward adaptability to a variety models (e.g., *in vivo*, *in vitro*). Plus, the rodent *in vivo* version (OECD TG 489 (OECD, 2016a)) has regulatory recognition. Adapting the Comet assay for use with human-derived 3D skin models for *in vitro* genotoxicity testing therefore, not only appears to be a straightforward choice for overcoming any species-specific differences but is also logical in terms of potential regulatory applicability as a component of tiered testing. The published literature included in the TST makes a compelling case for the RS Comet assay being sufficiently metabolically competent and thus a biologically fit for purpose assay - for non-systemic dermal genotoxicity testing. Furthermore, the air-liquid-interface (ALI) culture conditions of the RS Comet assay facilitates testing of a wide range of substances, including lipophilic substances at dose ranges relevant for topical skin exposure or application.

(b) Analysis of the regulatory rationale provided in the Test Submission

The regulatory rationale is described clearly in the TST. The need to develop reliable *in vitro* tests for genotoxicity testing associated with skin exposure was predominantly triggered by the European animal testing ban for cosmetics in conjunction with inherent challenges of alternative testing systems, one being the limited ability of existing cell-based *in vitro* methods to deliver sufficiently reliable predictions (e.g., elevated number of false positives). In addition, existing cell-based methods (i.e., MNT, CA, Ames test, and mammalian cell gene mutation tests) utilize rat liver S9 fraction to test for bioactivation of chemicals, which has an obvious limitation in the assessment of human skin genotoxicity. Citations provided in the TST (Pfuhler et al., 2020; Wiegand et al., 2014) along with those identified by the ESAC (Bataillon et al., 2019; Pfuhler et al., 2021) discuss the difference in metabolic enzyme expression in skin *vs.* liver and in rat *vs.* human, which provides a strong argument for the development a metabolically competent skin genotoxicity assay for assessment of human risk. These citations also provide evidence that the RS Comet assay uses a metabolically competent human skin model.

The RS Comet assay is intended to be used within regulatory genotoxicity hazard identification testing strategies to follow up positive results from the classical *in vitro* test battery, which is used as a first step. It is proposed to be used instead of *in vivo* genotoxicity test methods (OECD TG 489 (OECD, 2016a) in the context of test items that are causing genotoxicity in the form of DNA strand breaks *in vitro* and are primarily associated with dermal exposure.

The TST makes a clear case for the performance of the test method and for the potential use of the RS Comet assay as a standalone test for detecting genotoxicity associated with skin exposure. However, as presented, the TST does not address concerns associated with potential future applications. For example, the ESAC sees the potential for using the combination of the RS Comet and RSMN assays as a substitute for *in vivo* genotoxicity assays and questions whether this use scenario will sufficiently address potential mutagenicity, or systemic effects from skin penetration and absorption. Importantly, the testing strategy is only being evaluated by the ESAC for the context of

use, in which the RS Comet assay is used in combination with a RSMN evaluation for *in vitro* skin genotoxicity testing following topical exposure (see section 13).

1.3 Appraisal of the appropriateness of the study design

The TST compiles data from several published studies, thereafter, compiled to support the validation of the RS Comet assay. This, together with the protracted timeline associated with completing the validation study, introduces challenges such as protocol modifications, change of participating laboratories, and varying numbers of technical and biological replicates. While a more targeted and concise approach would have been preferable, the overall design seems adequate with respect to critical parameters, including study design, laboratory independence, blinded sampling and independent result evaluation, all being sufficiently fulfilled. Moreover, the submitted SOP provides clear guidance on experimental design, result evaluation and substance calls.

1.4 Appropriateness of the statistical evaluation

The statistical methods used in the evaluation of the validation study data are standard, well-vetted methodologies and therefore appropriate for the purpose of evaluating the assay. However, no power calculation to support the number of chemicals necessary to assess the Within-Laboratory Reproducibility (WLR) or the Between-Laboratory Reproducibility (BLR) has been performed.

2. Collection of existing data

2.1 Existing data used as reference data

The *in vitro* test battery data and *in vivo* reference data were collected from the literature and the EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database (Kirkland et al., 2016; Corvi et al., 2018; Madia et al., 2020).

The RS Comet assay data that were used for the validation/prediction model (PM) were produced during Phases 1 and 2 of the validation study. The results of Phase 1 were published by Reisinger et al. (2018). The results of Phase 2 were published by Pfuhler et al. (2020). There are earlier publications which present data from the RS Comet assay development (e.g., Reus et al., 2013), but these data were not used for the validation/PM. No data from other sources were used in the validation report.

2.2 Existing data used as <u>testing</u> data

Not applicable. All data were the result of prospective testing.

2.3 Search strategy for retrieving existing data

Genotoxicity data were accessed from published literature and the EURL ECVAM database (Kirkland et al., 2016; Corvi et al., 2018; Madia et al., 2020). However, the search strategy was not specified in the TST.

2.4 Selection criteria applied to existing data

The selection criteria were based on the EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database.

3. Quality aspects relating to data generated during the study

3.1 Quality assurance systems used when generating the data

It was stated in subsection 3.5 of the TST that the acquisition of the data for the assessment of the reproducibility and predictive capacity did not follow any formal quality assurance system. Full Good Laboratory Practice (GLP) compliance is not a requirement for test method development or validation. The Test Submitters agreed to follow the safeguards recommended by Balls et al. (1995). The ESAC agrees with the quality assurance measures taken in this validation study.

3.2 Quality check of the generated data prior to analysis

As noted in the TST, acquisition of the data for the assessment of the reproducibility and predictive capacity did not follow any formal quality assurance system. However, several safeguards recommended by Balls et al. (1995) were applied, which the ESAC considers acceptable.

4. Quality of data used for the purpose of the study (existing and newly generated)

4.1 Overall quality of the evaluated testing data (newly generated or existing)

The quality of the data provided is high. Data are properly reported and appropriate quality criteria for analysing the data were provided. All underlying data used for analyses were provided with the TST.

4.2 Quality of the reference data for evaluating relevance[1]

The Test Submitters used reference chemicals supported by the literature and documented in the EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database.

4.3 Sufficiency of the evaluated data in view of the study objective

The strategy for test chemicals selection has not been clearly described or justified relative to the context of use described in the TST or the supporting documents. A balanced set of TP, TN, and MP would seem to be more appropriate for a standalone use of the method rather than confirmation of positives obtained using traditional genotoxicity assays as was the proposed context of use. Nonetheless, the ESAC considers the quality of the RS Comet dataset sufficient to draw conclusions on assay validity (see subsections 4.1 and 4.2).

[1] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

5. Test definition (Module 1)

5.1 Quality and completeness of the overall test definition

Information provided regarding the biological and mechanistic relevance of the method is considered sufficient. When using the RS Comet assay, test items are assessed in a commercially available 3D RS model (Phenion® Full-Thickness Skin Model). The RS Comet assay is biologically relevant for investigating DNA damage in the skin after topical exposure.

A detailed SOP was provided by the Test Submitters (see Attachment 1.1 provided with the TST) together with supporting published papers, clearly describing the test system used, the endpoint measured, the quality criteria applied to the test system, the acceptance criteria applied to the results, and the positive and negative controls. Over the years, several modifications of the protocol have been introduced to optimise it, including single cell isolation, microscopical analysis, experimental design, PM, cytotoxicity assessment and DNA damage read-outs. The protocol recommended for future use is the version reported in Attachment 1.1 provided with the TST. The positive control in standard experiments is methyl methane sulfonate at a concentration of 5 $\mu g/cm^2$, dissolved in acetone. In experiments using the DNA repair inhibitor aphidicolin, which are required to confirm negative results (Brinkmann et al., 2013; Reisinger et al., 2018), the positive control is benzo(a)pyrene at a concentration of 12.5 $\mu g/cm^2$, dissolved in acetone. The RS Comet assay measures three readouts: one related to DNA damage and two cytotoxicity readouts (intracellular ATP and adenylate kinase leakage) after topical exposure over a period of 48 h. Test item concentrations exceeding those cut-off values for toxicity are not assessed for DNA damage.

For data analysis and data interpretation (paragraph 2.1.6, point of the TST), the Test Submitters have considered the procedure as applied for the *in vivo* study (OECD TG 489 (OECD 2016a)). Although two different PMs were explored during the validation study (Reisenger et al., 2018; Pfuhler et al., 2020), the data presented in the TST were analysed with one PM (PM1: an ANOVA followed by the one-sided Dunnett test) since it was stated by the Test Submitters that both PMs performed comparably well.

5.2 Quality and completeness of the documentation concerning protocols and prediction models

The SOP of the RS Comet assay evolved over several years and is supported by several publications. The SOP is well-described, including a description of the purpose, materials, study design, and assay procedure. It includes a clear data evaluation section which describes data processing, validity criteria, and evaluation of the results.

6. Test materials

6.1 Sufficiency of the number of evaluated test items in view of the study objective

The main selection criterion for the initial selection of chemicals appears to be based on matched data from *in vivo* rodent skin exposure studies. This approach yielded a limited list of test chemicals that was later expanded to enlarge the substance pool for statistical and validation purposes, as well as to include more data rich substances. Attachment 9 provided with the TST included information on basic physicochemical properties, structural information, and a full list of test substances. There were no statistical tests conducted at study initiation to justify the number of chemicals selected for the validation study. Nevertheless, the number of chemicals tested seems sufficient to assess the capacity of the method to detect expected modes of action and to evaluate reproducibility of the assay. The resulting list is fit for purpose in terms of the number of substances tested for assessing WLR and BLR. The ESAC noted that only a limited number of chemicals was selected for evaluating metabolic activation.

6.2 Representativeness of the test items with respect to applicability

The test items were selected for the validation study according to their description as TN, TP, and MP. The ESAC considers the test items sufficiently representative for the purpose of the study. Information was also provided on the genotoxic mode of action (e.g., alkylating agent). It should be noted that the TST does not make any statements related to defining the applicability domain or the representativeness of substance class/underlying mechanisms and coverage of technical parameters such as lipophilicity/solubility and stability.

7. Within-laboratory reproducibility (WLR) (Module 2)

7.1 Assessment of repeatability and reproducibility in the same laboratory

The WLR of the RS Comet assay was assessed during two phases of the validation study and is described in two separate publications (Reisinger et al., 2018; Pfuhler et al., 2020).

A total of 24 test items (8 from Phase 1 and 16 from Phase 2) were used to demonstrate WLR. While the number seems reasonable, no power calculation to determine the number of chemicals necessary to assess the WLR or the BLR has been performed. In addition, none of the 5 laboratories participating in the validation study tested all 24 chemicals (see Table 1).

The test items used in the validation study are listed in Attachment 4 provided with the TST together with their CAS numbers, commercial source, purity, physical form, and relevant physical/chemical properties. For those items, a range of toxic effects (toxic, non-toxic) and various modes of action are reported by the Test Submitters.

Results for WLR (excluding experiments which were deemed invalid due to protocol validity criteria, not expert judgement as described in Attachment 4 provided with the TST) are as follows with the caveat that most of calculations are based on data from only 2 experiments per test item (see Table 1).

Table 1. Reproducibility within each laboratory over time (WLR) in Phases 1 and 2

	Discordant	**Concordant**	**Total**	**%**
Lab A	0	7	7	**100**
Lab B	1	7	8	**87**
Lab C	0	3	3	**100**
Lab D	1	3	4	**75**
Lab E	0	3	3	**100**
All labs	2	23	25	**91**

Some potential sources of variability (e.g., test substance precipitation, background levels, dose selection) were described by the Test Submitters, but measures to control them were not proposed. Outlying values for cadmium chloride, 7,12-Dimethylbenz[a]-anthracene, and 2-acetylaminofluorene were identified from the results obtained in the WLR study. Reasons for these outlying values were discussed in the TST.

7.2 Conclusion on within-laboratory reproducibility as assessed by the study

Typically, three independent experiments are required for the assessment of WLR. For the RS Comet assay, most test items used by the Test Submitters to assess WLR have only been evaluated in two experiments. Only three chemicals had three replicates (i.e., 7,12-dimethylbenz[a]-anthracene and Di-(2-ethylhexyl)phthalate in Lab A and cyclohexanone in Lab B). Results are concordant in those three cases and exhibit high WLR based on three experiments. Given these results, the ESAC considers there is some uncertainty in the conclusion of the WLR based on two experiments only.

8. Transferability (Module 3)

8.1 Quality of design and analysis of the transfer phase

A detailed SOP has been provided as Attachment 1.1 to the TST. The SOP submitted clearly describes the test system used, the endpoint measured, the quality criteria applied to the test system, the acceptance criteria applied to the results, and the positive and negative controls.

Transferability was carried out using only one test item: methyl methane sulfonate, the positive control. All the laboratories participating in the validation study demonstrated successful transfer of the experimental procedures of the RS Comet assay in two parallel preceding projects using this chemical. The physical and chemical properties of methyl methane sulfonate and test results for this chemical were reported in Attachment 9 provided with the TST.

From the information provided by the Test Submitters, it appears that the RS Comet assay can be transferred to laboratories with competencies in the alkaline Comet assay in accordance with OECD TG 489 (OECD, 2016a). The Test Submitters claimed that standard technical laboratory skills and expertise in interpreting results of genotoxicity test methods are sufficient to conduct the RS Comet assay. No specific training is, therefore, recommended by the Test Submitters. However, it is clear from the associated publications (e.g., Pfuhler et al., 2020) that expert judgement is a critical aspect of results interpretation when using the RS Comet assay. For that reason, the ESAC recommends the use of this assay in laboratories experienced in the evaluation of Comet assay data (*in vivo* or *in vitro*) or to gain experience through testing of compounds with known activity in the Comet assay prior to testing unknown chemicals (i.e., Proficiency testing).

Recommendations are provided for critical aspects of the SOP (e.g., reagent purity, protecting Phenion® Full-Thickness Skin Model from direct UV light). Five consecutive experiments with methyl methane sulfonate or benzo(a)pyrene are suggested for a naïve laboratory with no previous experience in RS Comet assay.

Successful training should be demonstrated by testing at least one positive or, preferably, two positive and one negative test items not previously tested, following the entire protocol, i.e., from the determination of solubility to the final judgement of the result. The Test Submitters recommend selecting the test item(s) for training from the list of those tested in the validation exercise.

8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study

The SOP supporting the transfer to laboratories with competencies in alkaline Comet Assay is in accordance with OECD TG 489 (OECD, 2016a). Transferability was assessed with one test item (methyl methane sulfonate, the positive control). All five laboratories participating in the validation study demonstrated successful transfer in previous studies. The Test Submitters recommend demonstrating successful transfer of the test to all new laboratories by testing more than one positive test item with five consecutive experiments (or preferably one positive and one negative test item). The five experiments must meet the test validity criteria and must show clear increases in DNA migration.

The ESAC's opinion is that one test item is not enough to demonstrate transferability. More than one test item is required to increase confidence in the method, and this should include two positive and one negative test items.

9. Between-laboratory reproducibility (BLR) (Module 4)

9.1 Assessment of reproducibility in different laboratories

The BLR was assessed testing 8 test items (i.e., cadmium chloride, N-ethyl-N-nitrourea, 7,12-Dimethylbenz[a]-anthracene, propyl gallate, eugenol, Di-(2-ethylhexyl)phthalate, cyclohexanone, and mitomycin C) in five laboratories. See Attachments 8 and 9 provided with the TST for additional information.

The BLR of the RS Comet assay was assessed by the Test Submitters and reported to be equal to 87.5% (consistent with information presented in Attachment 8 provided with the TST).

Cadmium chloride did not produce reproducible positive responses. Given the well-known biochemistry of cadmium chloride (i.e., non-specific binding), it is not uncommon to obtain variable results.

Seven test items had concordant calls in three laboratories and one test item, cadmium chloride, was positive in two laboratories and negative in one. This resulted in a BLR estimate of **7/8 = 87.5%.** The results are summarised in Attachment 8 provided with the TST.

Potential experimental sources of variability that can result in reduced BLR primarily relate to experiments with borderline results and differences in dose selection as well as differences in chemical-cytotoxicity levels between laboratories. These factors may also explain why some substances are positive in main experiments performed in some laboratories and positive in experiments where aphidicolin was included to increase sensitivity in other laboratories, as observed for 7,12-dimethylbenz[a]-anthracene in the validation study.

9.2 Conclusion on between-laboratory reproducibility as assessed by the study

Typically, three independent laboratories are required for the assessment of BLR. In the RS Comet assay, five laboratories tested 8 test items (with at least 3 laboratories out of the 5 involved in each). Concordant results were obtained with 7 test items in three laboratories, whereas cadmium chloride provided discordant results in one laboratory versus the other two, with no obvious explanation provided. Based on these results, BLR was high (i.e., 87.5%).

10. Predictive capacity and overall relevance (Module 5)

10.1 Adequacy of the assessment of the predictive capacity in view of the purpose

Based on the materials provided by the Test Submitters, it is unclear if there is a Lead Lab.

Overall, a limited number of chemicals (i.e., 32 test items) were used to assess the predictive capacity of the RS Comet assay (Attachment 9 provided with the TST). Information on mode of action, reference *in vitro* and *in vivo* genotoxicity data, as well as carcinogenicity data are reported in Attachment 10 of the TST. The 32 chemicals comprised 8 liquids and 24 solids, which all had a low to very low vapour pressure. The chemicals were, except for cadmium chloride, organics representing various chemical classes and featuring numerous functional groups. To explore the potential of the test method to detect pro-mutagens, 3 pro-mutagenic chemicals (i.e., benzo(a)pyrene, 2-acetylaminofluorene (Lab B), and 2,4-diaminotoluene (Lab B)) were used in the development and optimisation of RS Comet assay. However, 7 pro-mutagenic chemicals were tested during the validation study (cyclophosphamide (Lab B), benzo(a)pyrene (Lab A), 7,12-dimethylbenz[a]-anthracene (Labs A, B, and C), 2-acetylaminofluorene (Lab B), 2,4-diaminotoluene (Lab B), 2-amino-3-methylimidazo[4,5-f]quinolone (Lab A), and 4-chloroaniline (Lab B)). Of the 15 test items used as TP, 7 test items needed metabolic activation to show positive results (Attachment 10 provided with the TST). Chemicals were initially selected by external experts from a master list prepared by Cosmetics Europe. As discussed above, test items cover different modes of action and a range of toxic effects.

The predictive capacity of the RS Comet assay was calculated based on 3 different scenarios (paragraph 2.5.2 of the TST). Data from 5 laboratories were combined, with a maximum of 18 chemicals tested in a single laboratory (Lab B). In the validation study, 8 test items (5 false negative and 3 false positive or equivocal prediction) were not correctly predicted in at least one laboratory.

Based on the PMs, e.g., for scenario A, sensitivity among the different laboratories ranged from 25% (Lab D) to 100% (Labs C and E) and specificity from 86% (Lab A) to 100% (Labs C, D, and E), with an overall accuracy (mean) > 83%. The low sensitivity of Lab D is due to the limited number of chemicals tested (n = 2). For Lab B, which tested the highest number of chemicals (i.e., 18): sensitivity was 80%, specificity 87%. The low sensitivity in Lab D may indicate a possible problem in the implementation of the method. The ESAC notes that only one chemical was used in the transfer phase, which may be not sufficient.

Only one TP that needed metabolic activation (i.e., DMBA) showed a consistent positive study outcome in three different laboratories. All other metabolic active TPs were tested in only one laboratory and showed either a positive study outcome (i.e., 2-AAF, BaP, and CP) or a negative study outcome (i.e., 2,4-DAT, 4-Chloroaniline, and IQ). These results suggest that not enough data is provided to conclude on the applicability of the RS Comet assay to test metabolic active genotoxic compounds.

The RS Comet assay, with its promising predictive capabilities, has potential utility for identification of genotoxic hazard. To that point, Lab B, which tested 18 chemicals, demonstrated a sensitivity and specificity higher that 80%. However, the validation study suffers from the lack of an a priori validation design that could have improved the quality of the study and the evaluation of predictive capacity.

10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

The overall relevance and accuracy of the RS Comet assay as a confirmation of a positive result in the Ames and/or mammalian cell gene mutation tests following dermal exposure is adequate.

11. Applicability domain (Module 6)

11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

For organizational purposes, the ESAC discusses the strengths and limitations of three different aspects of the applicability domain below.

Biological domain: As proposed in the TST, the RS Comet assay would be used for screening chemicals for potential genotoxicity resulting from dermal application of the chemical. The test was designed primarily for the purpose of screening cosmetic ingredients and is therefore focused on the relevant portal of entry (i.e., skin). Use of 3D RS allows for a more biologically relevant tissue architecture, which is more likely to recapitulate *in vivo* chemical penetration and toxicological response. Based on the evidence provided in the TST, the ESAC concurs that the assay is useful for the purpose of identifying chemicals with the potential to cause genotoxicity following dermal application – in a tiered testing paradigm that includes the standard *in vitro* genotoxicity assays and the RSMN assay as described in section 1. However, for the broader purpose of human health risk assessment, additional evidence would be required to make a regulatory decision, as this assay lacks the ability to account for potential systemic effects from dermal exposure.

Chemical domain: The applicability domain of the test items was not specifically explored in the TST. However, there are clear advantages to the use of a 3D RS model that is cultured at the ALI compared to traditional genotoxicity assays performed in submerged 2D monocultures, particularly in terms of the potential to increase the chemical applicability domain. The RS Comet assay utilizes a full-thickness skin model that also has the potential to be useful for identifying metabolically activated chemicals in a human-relevant context. This attribute is a significant advantage over immortalized or cancer cell line-based assays currently used in the traditional 2D submerged assays that rely on rat liver S9 fraction to approximate human skin metabolism.

Evidence of metabolic competence and ability of the RS Comet assay to identify bioactivated genotoxicants is provided in the TST and referenced publications (e.g., Reisinger et al., 2018; Pfuhler et al., 2020; Brinkmann et al., 2013; Reus et al., 2013). Out of 16 test items used as TP chemicals, 7 that require bioactivation were evaluated during the validation study. Of those, 4 were correctly identified as mutagenic (i.e., benzo(a)pyrene, cyclophosphamide, 7,12-Dimethylbenz[a]-anthracene, and 2-acetylaminofluorene) whereas the remaining three were not (i.e., 2-amino-3-methylimidazo[4,5-f]quinoline, 2,4-Diaminotoluene, and 4-chloroaniline). However, it should be noted that benzo(a)pyrene (Brinkmann et al., 2013) and 2,4-diaminotoluene (Reus et al., 2013) were correctly classified during assay development and optimization while the results for 2-acetylaminofluorene were not published or included in the TST. In addition, because the model involves culturing at the ALI, a broader universe of chemicals and substances that are not generally well-tolerated by submerged *in vitro* assays can be tested (e.g., lipophilic and otherwise poorly dissolved in cell culture medium). Furthermore, complex mixtures and formulations in various vehicles (creams, emulsions, etc.) could potentially be tested in a skin model cultured at the ALI. Since the TST did not specifically evaluate lipophilic chemicals or product formulations, no conclusions can be drawn on the utility of this assay for such purposes. However, the design of the assay suggests the potential to expand beyond the limited chemical space evaluated in the validation study, which defines all submerged 2D *in vitro* systems.

Mechanistic domain: Regarding the chemical space covered by the chemicals evaluated in the validation study, the TST would have benefitted from additional details regarding underlying mechanisms, substance class representation, and test chemical stability. Hence, there remain some uncertainties regarding the reasoning of some of the substance selections. A recently developed AOP for genotoxicity (Sasaki et al., 2020) provides a streamlined, science-backed series of key events for various mechanisms of chemical-induced genotoxicity and would be a valuable resource for better defining the domain of applicability for this assay in terms of mode of action.

In conclusion, only a very limited number of chemicals have been tested (i.e., n = 32) and information related to the genotoxic mode of action provided in the TST was also limited. For that reason, the applicability domain of the RS Comet assay will need to be evaluated further along with its potential for broader use.

11.2 Quality of the description of applicability domain, limitations, exclusions

This study did not attempt to define the applicability domain in terms of chemical space. As such, this cannot be evaluated by the ESAC.

12. Performance standards (Module 7)

12.1 Adequacy of the proposed Essential Test Method Components

Not applicable.

12.2 Adequacy of the proposed Reference Chemicals

Not applicable.

12.3 Adequacy of the proposed performance target values

Not applicable.

13. Readiness for standardised use

13.1 Assessment of the readiness for regulatory purposes

The RS Comet assay should be considered together with the RSMN assay, as they are intended to be used within tiered regulatory genotoxicity hazard identification testing strategies to follow up positive results from the classical *in vitro* test battery. The RS Comet assay is proposed to be used instead of *in vivo* genotoxicity test methods (e.g., OECD TG 489 (OECD, 2016a)), in the context of test items that are causing DNA strand breaks *in vitro* and are primarily associated with the dermal exposure route. In combination with the RSMN assay, they cover all genotoxicity endpoints that usually need to be addressed for regulatory purposes (gene mutation, clastogenicity, and aneugenicity).

The approach proposed by the Test Submitters involves conducting the RS Comet assay as a follow-up to a positive Ames and/or mammalian cell gene mutation tests and a negative result in the RS Comet assay would support overriding the initial positive result. Similarly, a positive *in vitro* MNT or CA would trigger a follow-up with the RSMN assay, and a negative result in the RSMN would support overriding the initial positive result. According to the Test Submitters, in cases where both Ames/ mammalian cell gene mutation tests and MNT/CA studies yield positive results, follow-up testing should be conducted with both the RS Comet and RSMN assays, and a positive follow-up result in either the RS Comet assay or the RSMN assay would lead to the conclusion of a genotoxicity hazard associated with topical exposure.

As a precautionary note, test items requiring metabolic activation may be classified as false negatives based on RS genotoxicity tests alone. This is particularly crucial when these compounds are accessible for systemic absorption. Considering the skin's lower metabolic capacity, initial results might appear negative. However, once absorbed, chemicals could undergo systemic bioactivation, thereby elevating the risk of neoplasia, which is an unacceptable outcome.

The 12th Revision of the Scientific Committee on Consumer Safety (SCCS) Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation, concluded that (Pfuhler et al., 2020; Pfuhler et al., 2021; SCCS, 2023):

a. 3D tissue models simulate *in vivo*-like conditions including cell viability, proliferation, differentiation, morphology, gene and protein expression. These models can complement classical 2D cell culture-based assays;
b. 3D tissue-based genotoxicity assays can be used as 2nd tier assays to follow-up on positive results from standard *in vitro* assays;
c. For adoption of a tissue model as a 2nd tier assay, ability to detect the full range of genotoxic damage (leading to mutagenicity, clastogenicity, aneugenicity) should be demonstrated;
d. The 72-h protocol for the RSMN has higher sensitivity than the 48-h protocol;
e. The RS Comet assay (Pfuhler et al., 2020) and RSMN (Pfuhler et al., 2021) assays are now sufficiently validated to move towards individual OECD TGs, but an independent peer review of the validation study is still needed.

13.2 Assessment of the readiness for other uses

For the intended purpose, no mechanistic limitations of the RS Comet assay are known. Since only a very limited number of chemicals have been tested (n = 32), its applicability for screening purposes will need to be evaluated further. The applicability to mixtures and final formulations has not been demonstrated yet.

13.3 Critical aspects impacting on standardised use

In the validation study, which was conducted over a protracted period of time, the composition of the participating laboratories changed. While results are promising, the low sensitivity (i.e., 25%) obtained in Lab D should not be minimised as it may indicate possible problems in the implementation of the method. In the transfer phase only one chemical was used, which may be not sufficient.

The method has been optimized and it offers several advantages compared to traditional 2D submerged assays and no critical deficiencies have been identified. The only technical limitation identified relates to test item solubility. The suitability of solvents other than acetone and 70% ethanol, has not been tested.

13.4 Gap analysis

Even if different chemical classes have been tested, due to a small number of chemicals tested (n = 32), the applicability domain of the RS Comet assay will need to be evaluated further.

14. Other considerations

The ESAC sees value in potential expanded use for this method and the combination with the RSMN for the evaluation of other product categories and configurations (e.g., pesticides, chemical mixtures, final products), provided systemic bioavailability is considered.

The RS Comet and RSMN assays were designed for the follow up of a positive Ames and/or mammalian cell gene mutation tests, and/or a positive MNT and/or CA performed as part of the traditional *in vitro* genotoxicity test battery, respectively. As such, the assays can aid with weight-of-evidence (WoE)-based regulatory decision-making for potentially genotoxic substances with skin contact. In this context, the use of 3D RS models not only offers the advantage of organ similarity with regard to barrier function but also (limited) metabolic competence. Unlike other test systems, however, the models also provide human relevance. It should be noted that, as presented by the Test Submitters, the system will not allow any predictions on potential systemic genotoxicity. Accounting for systemic genotoxicity will require additional considerations such as the use of liver S9 fraction and/or data on skin absorption/penetration. In cases of significant skin absorption of a compound found to be positive in traditional genotoxicity assays following metabolic activation (by S9) but negative in the RS assays, the possibility of its liver activation into a genotoxic compound cannot be excluded due to potential differences in the metabolic capacity of the skin and the liver. Hazard classification should be straight forward for all substances identified as initial positives which are not subject to S9-activation as these will either be confirmed or overruled by subsequent testing with RS assays. However, for substances originally requiring S9-activation, overruling in the RS assays is less straight forward as a negative result obtained using the RS assays cannot principally exclude the possibility of systemic activation following skin penetration. Therefore, such substances will have to undergo further testing followed by WoE analysis.

15. Conclusions on the study

15.1 ESAC SG summary of the results and conclusions of the study

General comments

The ESAC agrees with the proposed use of the RS Comet assay as a confirmatory assay to follow up positive *in vitro* Ames and/or mammalian cell gene mutation test findings following dermal application for potential genotoxic chemicals.

The RS Comet is not currently proposed as a stand-alone assay for systemic genotoxicity. If it were to be considered for use in systemic genotoxicity assessment, data derived from the assay would need to be considered together with other data in a WoE approach.

The ESAC had numerous discussions on the role of this assay in a WoE approach. A demonstration of some of the considerations that the ESAC believes are important are discussed in section 14 above. These considerations include, but are not limited to, the degree of metabolic activation and potential systemic availability.

The ESAC sees value in potential expanded use for this method and the combination with the RSMN assay. See other considerations (section 14) for additional discussion.

Chemicals selection

The strategy for chemicals selection has not been clearly described or justified relative to the context of use described in the TST or in the supporting documents. To this point, there appears to be a disconnect between the proposed purpose of the assay (i.e., confirmation of positive results obtained in the standard *in vitro* genotoxicity battery) and the chemicals selection strategy. A balanced set of TP, TN, and MP substances would seem to be more appropriate for a standalone use of the methods rather than the proposed use of confirming positive results. The ESAC also notes that the test items were selected and classified as TP, TN, and MP by subject matter experts that were not part of the validation study and that the rational for those decisions was not provided in the TST. Consequently, there is some uncertainty associated with these classifications that the ESAC could not resolve.

Validation study design

The TST compiles data from several published studies, thereafter, compiled to support the validation of the RS Comet assay. This, together with the protracted timeline for completing the validation study, introduce challenges such as protocol modifications, change of participating laboratories, and varying numbers of technical and biological replicates that were tested in the participating laboratories. While a more targeted and concise approach to validation would have been preferable, the overall design seems adequate with respect to critical parameters such as study design, laboratory independence, blinded sampling, and independent result evaluation. Moreover, the submitted SOP provides clear guidance on experimental design, result evaluation and substance calls.

Conclusions on the chemical space of the RS Comet assay cannot be made due to the limited number of chemicals tested and since the test substance applicability domain was not specifically explored in the TST. For example, the TST did not specifically evaluate lipophilic chemicals. It is noted that the RS Comet assay has the potential to be useful for identifying metabolically activated chemicals in a human relevant context because three of five chemicals requiring metabolic activation were correctly identified as genotoxic by the RS Comet assay (Reisenger et al., 2018; Pfuhler et al., 2020). While that is the case, it would have been preferrable to have tested the seven TP test items requiring metabolic activation in more than one laboratory to show consistency between laboratories. With the data currently provided, ESAC cannot conclude on the potential of the RS Comet assay to identify metabolically activated chemicals.

Overall, the BLR is acceptable. However, the ESAC has some concerns with respect to the transferability and the WLR. The ESAC's opinion is that one test item is not enough to demonstrate transferability. More than one test item is required to increase confidence in the method, and this should include two positive and one negative test items. With the evidence for BLR, there may be enough evidence for transferability. Furthermore, because there were only two experiments, there is some uncertainty in the conclusion of the WLR. The overall relevance and accuracy of the assay as a confirmation of a positive result in the Ames or mammalian cell gene mutation tests is adequate.

Despite these limitations, the ESAC concludes that the results of the validation study regarding reproducibility, transferability and predictive capacity are adequate for the defined purpose.

Biological, chemical and mechanistic domain of the assay

Compared to traditional genotoxicity assays performed in submerged 2D monocultures, there are clear advantages to the use of the RS Comet assay, which is composed of a 3D skin model cultured at the ALI. Because the model possesses *in vivo*-like tissue structure, more types of chemicals can be tested. While that is the case, additional studies are needed to fully define the applicability domain, including the metabolic competence of the RS Comet assay by testing more TP test items that require metabolic activation and testing them in more than one laboratory. Based on the data available, neither the RSMN assay (metabolic competence may be limited due to maturation state of keratinocytes) nor the RS Comet assay (not enough data to conclude on potential to correctly classify metabolic active test items) demonstrated success identifying metabolically activated genotoxicants. However, the combination of the RS Comet and RSMN assays correctly identified six out of seven TP test items that require bioactivation (strategic combination of 1). The only outlier (i.e., diaminotoluene) was negative in the RS Comet and positive in 1 out of 3 laboratories in the RSMN (strategic combination of 0.33). Therefore, also for bioactive chemicals, using the combination of the RSMN and RS Comet assays is preferred.

15.2 Extent to which study conclusions are justified by the study results alone

Data presented in this study were published in peer reviewed publications. The RS Comet assay was designed primarily for the purpose of screening cosmetic ingredients and is therefore focused on portal of entry, specifically skin. However, laboratory transferability was, depending upon the laboratory, demonstrated using only one chemical during the transfer phase. Consequently, there is some doubt about whether the method was successfully transferred to all participating laboratories, which may affect the reproducibility and predictive capacity of the model. Rather than testing all chemicals in three laboratories, the Test Submitters applied a lean design that included more laboratories and the testing of relatively few chemicals in each laboratory. Consequently, the uncertainty around the values obtained from any one laboratory is high. However, when all data are combined the calculated WLR is 91% and the study conclusions are supported by the data. Similarly, with only a limited number of chemicals included in the design, BLR of 87.5% was achieved. Despite some deficiencies in the study design (as noted above), the ESAC concludes that WLR and BLR are sufficient and justified by the study results.

15.3 Extent to which conclusions are plausible in the context of existing information

All published data generated using the RS Comet assay have been included in the TST. Additionally, the ESAC is aware that the RS Comet assay has been used to generate data used to support the regulatory assessment of cosmetic ingredients.

16. Recommendations

16.1 General recommendations

The validation study would have benefited from applying a more comprehensive, a priori validation strategy. The ESAC sees validation as a process of continuous confidence building rather than a single ring trial study. To that point, the ESAC does not see a problem with how this study was conducted (i.e., multiple steps). It is, however, critical that studies like the one evaluated here are coherent across study phases to promote confidence in the data. Specific recommendations related to the validation and application of the RS Comet assay include:

- Apply a more mechanistically driven approach for chemicals selection. To ensure sufficient coverage of the chemical space for the intended purpose of the method, it would have been helpful if the chemicals selected for the validation were coordinated at study initiation to evaluate, for example, metabolism and *in vitro* biokinetics.
- The ESAC suggests demonstrating method transferability by testing one negative and two positive test items (as opposed to relying solely on a single test item).
- Predictive capacity was determined using a relatively small set of chemicals (i.e., 32). Including more chemicals in the validation would have been preferred.
- As a best practice, three replicate experiments within a laboratory should be used for a validation study to assess WLR.
- As a best practice, at least three laboratories should be used for a validation study to assess BLR.
- Due to the limited number of test items evaluated, additional studies are needed to fully define the applicability domain. To that point, the ESAC recommends further evaluating the utility of the RS Comet assay, including more compounds that require bioactivation. This could be accomplished by using more publicly available data as it becomes available.
- The suitability of solvents for use in the RS Comet assay other than acetone and 70% ethanol should be investigated and guidance for selection of solvents for specific test items should be included in the SOP.
- Future applications:
    - The assay is fit for the application described in the TST. Because the method provides valuable information for compounds applied topically to the skin in general, the use application could potentially be broadened and the ESAC would recommend validation for other topically applied product classes such as mixtures, and formulations.
    - Due to differences in skin and liver metabolism, the ESAC expresses some concern related to the possibility that a topically applied chemical that penetrates the skin may become systemically available and could be bioactivated by the liver. End users cannot always assume that a compound applied topically to the skin will not be systemically available and metabolized by the liver. Hazard classification should be straight forward for all substances identified as initial positives which are not subject to S9-activation as these positive results will be either confirmed or overruled by subsequent RS-testing. However, for substances originally requiring S9-activation, overruling by the RS assays is less straight forward as a negative RS-result cannot principally exclude the possibility of systemic activation following skin penetration. Therefore, such substances will have to undergo further testing followed by WoE analysis (see other considerations, section 14).

16.2 Specific recommendations (e.g., concerning improvement of protocols)

Not applicable.

17. References

- Balls M, Blaauboer BJ, Fentem J, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, Reinhardt CA, Repetto G, Sladowski D, Spielmann H, Zucco F (1995) Practical aspects of the validation of toxicity test procedures: the report and recommendations of ECVAM workshop 5. Altern Lab Anim. 23(1):129-147. doi: 10.1177/026119299502300116. https://journals.sagepub.com/doi/10.1177/026119299502300116.
- Bataillon M, Lelièvre D, Chapuis A, Thillou F, Autourde JB, Durand S, Boyera N, Rigaudeau AS, Besné I, Pellevoisin C (2019) Characterization of a new reconstructed full thickness skin model, T-Skin™, and its application for investigations of anti-aging compounds. Int J Mol Sci 20(9):2240. doi: 10.3390/ijms20092240. https://www.mdpi.com/1422-0067/20/9/2240.
- Brinkmann J, Stolpmann K, Trappe S, Otter T, Genkinger D, Bock U, Liebsch M, Henkler F, Hutzler C, Luch A (2013) Metabolically competent human skin models: activation and genotoxicity of benzo[a]pyrene. Toxicol Sci 131(2):351-9. doi: 10.1093/toxsci/kfs316. https://academic.oup.com/toxsci/article/131/2/351/1643026?login=true.
- Corvi R, Madia F (2018) EURL ECVAM genotoxicity and carcinogenicity consolidated database of Ames positive chemicals. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames.
- Kirkland D, Speit G (2008) Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing *in vivo*. Mutat Res 654(2):114-32. doi: 10.1016/j.mrgentox.2008.05.002. https://www.sciencedirect.com/science/article/pii/S1383571805000835?via%3Dihub.
- Kirkland D, Kasper P, Martus HJ, Mueller L, van Benthem J, Madia F, Corvi R (2016) Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. Mutat Res 795:7-30. doi: 10.1016/j.mrgentox.2015.10.006.
- Madia F, Corvi R (2020) EURL ECVAM genotoxicity and carcinogenicity consolidated database of Ames negative chemicals. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/38701804-bc00-43c1-8af1-fe2d5265e8d7.
- OECD (2016a) TG 489: *In Vivo* Mammalian Alkaline Comet Assay, adopted 29 July 2016. https://www.oecd-ilibrary.org/environment/test-no-489-in-vivo-mammalian-alkaline-comet-assay_9789264264885-en.
- OECD (2016b) TG 474: Mammalian Erythrocyte Micronucleus Test, adopted 29 July 2016. https://www.oecd-ilibrary.org/environment/test-no-474-mammalian-erythrocyte-micronucleus-test_9789264264762-en.
- OECD (2022) TG 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays, adopted 30 June 2022. https://www.oecd-ilibrary.org/environment/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_9789264203907-en.
- Pfuhler S, van Benthem J, Curren R, Doak SH, Dusinska M, Hayashi M, Heflich RH, Kidd D, Kirkland D, Luan Y, Ouedraogo G, Reisinger K, Sofuni T, van Acker F, Yang Y, Corvi R (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 Genet Toxicol Environ Mutagen 850-851:503135. doi: 10.1016/j.mrgentox.2020.503135. https://www.sciencedirect.com/science/article/pii/S138357182030005X?via%3Dihub.

- Pfuhler S, Pirow R, Downs TR, Haase A, Hewitt N, Luch A, Merkel M, Petrick C, Said A, Schäfer-Korting M, Reisinger K (2021) 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 36(1):19-35. doi: 10.1093/mutage/geaa009. https://academic.oup.com/mutage/article/36/1/19/5802010.

- Reisinger K, Blatz V, Brinkmann J, Downs TR, Fischer A, Henkler F, Hoffmann S, Krul C, Liebsch M, Luch A, Pirow R, Reus AA, Schulz M, Pfuhler S (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. doi: 10.1016/j.mrgentox.2018.01.003. https://www.sciencedirect.com/science/article/pii/S1383571817302176?via%3Dihub.

- Reus AA, Reisinger K, Downs TR, Carr GJ, Zeller A, Corvi R, Krul CA, Pfuhler S (2013) Comet assay in reconstructed 3D human epidermal skin models--investigation of intra- and inter-laboratory reproducibility with coded chemicals. Mutagenesis 28(6):709-20. doi: 10.1093/mutage/get051. https://pubmed.ncbi.nlm.nih.gov/24150594/.

- Sasaki JC, Allemang A, Bryce SM, Custer L, Dearfield KL, Dietz Y, Elhajouji A, Escobar PA, Fornace AJ Jr, Froetschl R, Galloway S, Hemmann U, Hendriks G, Li HH, Luijten M, Ouedraogo G, Peel L, Pfuhler S, Roberts DJ, Thybaud V, van Benthem J, Yauk CL, Schuler M (2020) Application of the adverse outcome pathway framework to genotoxic modes of action. Environ Mol Mutagen 61(1):114-134. doi: 10.1002/em.22339. https://onlinelibrary.wiley.com/doi/10.1002/em.22339.

- SCCS (2023) Scientific Committee on Consumer Safety. Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation - 12th revision, from pages 74-75. Adopted May 15, 2023. SCCS/1647/22. https://health.ec.europa.eu/publications/sccs-notes-guidance-testing-cosmetic-ingredients-and-their-safety-evaluation-12th-revision_en.

- 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. doi: 10.1159/000358272. https://karger.com/spp/article/27/5/263/295819/Dermal-Xenobiotic-Metabolism-A-Comparison-between.

Annex 3

ESAC SUB-GROUP REPORT

ON RSMN

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health and Food
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(**ESAC**)

ESAC SUB-GROUP REPORT

on the

Scientific Validity of the RSMN Test Method

Title page information	
File name	ESAC_SG_Report_RSMN.docx
Abbreviated title of ESAC request	RSCOMET-RSMN
Relating to ESAC REQUEST Nr.	2023-01
Request discussed through	ESAC 48 (February 2023)
Report to be handed over to ESAC Chair and EURL ECVAM Coordinator by	Rebecca Clewell (Sub-Group Chair)

Version tracking

Date	**Version**	**Author(s)**	**Description**
18/12/2023	V1.0	ESAC SG	First agreed draft of ESAC SG Report
23/05/2024	V2.0	ESAC SG	Final approved draft of ESAC SG Report sent to ESAC for endorsement
24/01/2025	V3.0	ESAC	Final version of ESAC SG Report endorsed by the ESAC

Table of Contents

TABLE OF CONTENTS 42

ESAC SUB-GROUP (SG) 44

ABBREVIATIONS USED IN THE DOCUMENT 45

1. STUDY OBJECTIVE AND DESIGN 46

1.1 Analysis of the clarity of the study objective's definition 46

(a) ESAC SG summary of the study objective as outlined in the Test Submission 46

(b) Appraisal of clarity of study objective as outlined in the Test Submission 46

1.2 Quality of the background provided concerning the purpose of the test method 47

(a) Analysis of the scientific rationale provided in the Test Submission 47

(b) Analysis of the regulatory rationale provided in the Test Submission 48

1.3 Appraisal of the appropriateness of the study design 49

1.4 Appropriateness of the statistical evaluation 49

2. COLLECTION OF EXISTING DATA 50

2.1 Existing data used as reference data 50

2.2 Existing data used as testing data 50

2.3 Search strategy for retrieving existing data 50

2.4 Selection criteria applied to existing data 50

3. QUALITY ASPECTS RELATING TO DATA GENERATED DURING THE STUDY 51

3.1 Quality assurance systems used when generating the data 51

3.2 Quality check of the generated data prior to analysis 51

4. QUALITY OF DATA USED FOR THE PURPOSE OF THE STUDY (EXISTING AND NEWLY GENERATED) 52

4.1 Overall quality of the evaluated testing data (newly generated or existing) 52

4.2 Quality of the reference data for evaluating relevance 52

4.3 Sufficiency of the evaluated data in view of the study objective 52

5. TEST DEFINITION (MODULE 1) 53

5.1 Quality and completeness of the overall test definition 53

5.2 Quality and completeness of the documentation concerning protocols and prediction models 53

6. TEST MATERIALS 54

6.1 Sufficiency of the number of evaluated test items in view of the study objective 54

6.2 Representativeness of the test items with respect to applicability 54

7. WITHIN-LABORATORY REPRODUCIBILITY (WLR) (MODULE 2) 55

7.1 Assessment of repeatability and reproducibility in the same laboratory 55

7.2 Conclusion on within-laboratory reproducibility as assessed by the study 57

8. TRANSFERABILITY (MODULE 3) 58

8.1 Quality of design and analysis of the transfer phase 58

8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study 58

9. BETWEEN-LABORATORY REPRODUCIBILITY (BLR) (MODULE 4) 59

9.1 Assessment of reproducibility in different laboratories 59

9.2 Conclusion on between-laboratory reproducibility as assessed by the study 59

10. PREDICTIVE CAPACITY AND OVERALL RELEVANCE (MODULE 5) 60

10.1 Adequacy of the assessment of the predictive capacity in view of the purpose 60

10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose 61

11. APPLICABILITY DOMAIN (MODULE 6) **62**
11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions 62
11.2 Quality of the description of applicability domain, limitations, exclusions 63
12. PERFORMANCE STANDARDS (MODULE 7) **64**
12.1 Adequacy of the proposed Essential Test Method Components 64
12.2 Adequacy of the proposed Reference Chemicals 64
12.3 Adequacy of the proposed performance target values 64
13. READINESS FOR STANDARDISED USE **65**
13.1 Assessment of the readiness for regulatory purposes 65
13.2 Assessment of the readiness for other uses 65
13.3 Critical aspects impacting on standardised use 66
13.4 Gap analysis 66
14. OTHER CONSIDERATIONS **67**
15. CONCLUSIONS ON THE STUDY **68**
15.1 ESAC SG summary of the results and conclusions of the study 68
15.2 Extent to which study conclusions are justified by the study results alone 70
15.3 Extent to which conclusions are plausible in the context of existing information 70
16. RECOMMENDATIONS **71**
16.1 General recommendations 71
16.2 Specific recommendations (e.g., concerning improvement of protocols) 71
17. REFERENCES **72**

ESAC Sub-Group (SG)

Full title: ESAC Sub-Group on RS Comet and RSMN Assays

Abbreviated title: ESAC SG RSCOMET-RSMN

The European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) Scientific Advisory Committee (ESAC) Sub-Group (SG) was established in February 2023 during the 48th plenary meeting of the ESAC (ESAC48) to assist in the production of an ESAC Opinion on the scientific validity of the Reconstructed human Skin (RS) Comet and Micronucleus (MN) *in vitro* test methods for the assessment of genotoxicity.

This ESAC SG Report on the scientific validity of the RSMN test method was prepared at the request of EURL ECVAM by the "ESAC Sub-Group on RS Comet and RSMN Assays" (ESAC SG), which was charged with conducting a detailed scientific peer review of the external validation studies of the RS Comet and RSMN *in vitro* test methods. The basis for the scientific peer review was the EURL ECVAM Request for ESAC Advice approved by the ESAC during the ESAC48 plenary meeting of February 2023 (ESAC request 2023-01).

The ESAC SG met at the Joint Research Centre (JRC) in Ispra on 22-23/06/2023 and virtually on 13 and 26/10/2023; 8, 16, 20 and 24/11/2023; 1, 4, 11 and 18/12/2023; 10 and 22/01/2024; 5 and 14/02/2024; 1 and 22/03/2024; 3 and 6/05/2024, to conduct its peer review. The ESAC SG Report on RSMN was endorsed by the ESAC SG on 23/05/2024 and represents its consensus view. The Report was slightly revised and finally endorsed by the ESAC on 24/01/2025 following meetings on 8 and 20/01/2025.

The ESAC SG had the following members:

- Dr. Rebecca CLEWELL (SG Chair)
- Prof. Emanuela CORSINI
- Dr. Anne KIENHUIS
- Dr. David M. LEHMANN
- Prof. Blanca RODRÍGUEZ
- Prof. Helmut SEGNER
- Dr. Tewes TRALAU

EURL ECVAM (Secretariat):

- Dr. João BARROSO (ESAC Coordinator)
- Dr. Raffaella CORVI
- Dr. Milena MENNECOZZI

ABBREVIATIONS USED IN THE DOCUMENT

- **2D** Two-dimensional
- **3D** Three-dimensional
- **BLR** Between-laboratory reproducibility
- **CA** Chromosomal aberration test
- **ESAC** EURL ECVAM Scientific Advisory Committee
- **ESAC SG** ESAC Sub-Group
- **EURL ECVAM** European Union Reference Laboratory for Alternatives to Animal Testing
- **JRC** Joint Research Centre
- **MN** Micronucleus
- **MNT** Micronucleus Test
- **MP** Misleading Positive
- **OECD** Organisation for Economic Co-operation and Development
- **REACH** Registration, Evaluation, Authorisation and Restriction of Chemicals
- **RhE** Reconstructed human Epidermis
- **RS** Reconstructed human Skin
- **SCCS** Scientific Committee on Consumer Safety
- **SG** Sub-Group
- **SOP** Standard Operating Procedure
- **TG** Test Guideline
- **TN** True Negative
- **TP** True Positive
- **TST** Test Submission Template (the Full Test Submission)
- **Vs** Versus
- **WLR** Within-laboratory reproducibility
- **WoE** Weight-of-Evidence

1. Study objective and design

1.1 Analysis of the clarity of the study objective's definition

(a) ESAC SG summary of the study objective as outlined in the Test Submission

The Reconstructed human Skin Micronucleus (RSMN) assay is an *in vitro* genotoxicity assay for the detection of micronuclei. Micronuclei may contain chromosome fragments produced from DNA breakage (clastogens) or whole chromosomes produced by disruption of the mitotic apparatus (aneugens). Unless fully repaired, the respective DNA damage might either result in DNA mutation or chromosomal aberrations, and ultimately, carcinogenicity. Clastogenicity and aneugenicity are crucial endpoints for the toxicological safety assessment of substances.

Genotoxicity testing usually follows a tiered approach with higher tier testing routinely being conducted *in vivo* as a follow up of *in vitro* positive results. This approach creates challenges for substances where *in vivo* testing is banned, such as cosmetic ingredients. It also limits testing efficiency, negatively affects animal welfare, and fails to address societal expectations for more humane testing. The RSMN assay is an adaptation of the *in vivo* and the *in vitro* micronucleus tests (MNTs), which have OECD Test Guidelines (TGs) (OECD TG 474 (OECD, 2016a) and OECD TG 487 (OECD, 2023)). The RSMN assay transfers the micronucleus detection method to three-dimensional (3D) Reconstructed human Skin (RS) models (i.e., EpiDerm™ Reconstructed human Epidermis (RhE) model).

The RSMN Test Submission Template (TST) presents the assay as part of a tiered testing strategy for the evaluation of genotoxicity following dermal exposure. The respective strategy proposes the use of the RSMN assay as a follow up to confirm or reject a positive *in vitro* MNT and/or chromosomal aberration test (CA), and the RS Comet (see Annex 2: RS Comet Sub-Group Report) as a follow up to confirm or reject a positive Ames and/or mammalian cell gene mutation tests. While the proposed testing battery will not inform the underlying mechanisms of genotoxicity or possible substance or metabolite absorption, it nevertheless covers all possible types of DNA damage (i.e., gene mutations, clastogencity and aneugenicity). The RSMN and RS Comet assays both rely on 3D RS models with the proposed strategy aimed at regulatory use in the context of cosmetic ingredient testing for genotoxicity associated with skin exposure. However, the tests also have potential utility for the testing of pesticides, substances under Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), or dermally applied pharmaceutical products, applied topically. Herein, the ESAC provides a review of both the RS Comet and RSMN TSTs.

For chemicals in which dermal genotoxicity is the toxicological endpoint of concern following application to the skin, the ESAC recommends using both assays as a follow up to any positive result in the standard genotoxicity *in vitro* test battery. A positive result in either the RSMN or RS Comet would be sufficient to conclude that the test item is genotoxic. The ESAC recommends starting with the RS assay covering the mode of action that tested positive in the standard *in vitro* test battery and only performing the second RS assay in case of a negative result in the first RS assay.

(b) Appraisal of clarity of study objective as outlined in the Test Submission

The review of the RSMN assay by the ESAC relies on data collected and compiled over a period of more than 10 years. This approach brings inherent challenges, some of which are clearly reflected in the TST. For example, the remit of the validation study, as well as its design, were subject to several adjustments over time (see below).

The Test Submitters were able to retrospectively address most of these issues, although some uncertainties remain about the reasoning behind some of the test substance selections as well as the justification for some of the protocol adjustments. Regarding the chemical space covered in the validation studies, the TST would have benefitted from additional details related to the underlying mechanisms, substance class representation, and stability of substances selected for the validation

study. For example, only a small number of chemicals included in the validation study require bioactivation (i.e., 7 evaluated in Pfuhler et al., 2021). Data on 2 chemicals requiring metabolic activation, benzo[*a*]pyrene (BaP) and cyclophosphamide (CPA), not considered in this validation study, were evaluated by Kidd et al. (2021). Furthermore, the strategy for chemicals selection has not been clearly described or justified relative to the context of use described in the TST or in the supporting documents. To this point, there appears to be a disconnect between the proposed purpose of the assay (i.e., confirmation of positive results obtained in the standard *in vitro* genotoxicity battery) and the chemicals selection strategy. A balanced set of true positive (TP), true negative (TN), and misleading positive (MP) substances would seem to be more appropriate for a standalone use of the methods rather than the proposed use of confirming positive results. These limitations are addressed in later sections of this report (see subsections 6.2 and 11.2).

Nevertheless, the TST is clear about its general objective and contains the necessary controls and data, the majority of which is published and publicly available, to facilitate review by the ESAC. Further, while a more context-specific test item set would be preferred, the substances tested during the validation study included TP, TN, and MP. Likewise, metabolic competency was tested using a selection of chemicals that require bioactivation. Thus, the TST contains sufficient information for evaluation of the RSMN assay and clearly outlines the potential regulatory benefit of its application.

1.2 Quality of the background provided concerning the purpose of the test method

The TST adequately describes the purpose of the assay (i.e., follow-up test for substances testing positive in the *in vitro* MNT and/or *in vitro* CA, and for which genotoxicity was the toxicological endpoint of concern following application to the skin). The method design concept involved merging conventional genotoxicity parameters with existing 3D, RS technology. This integration led to the development of the "3D" RSMN assay, discussed in more detail here, and an RS Comet assay (see Annex 2: RS Comet Sub-Group Report). The combined use of these assays is considered suitable for investigating positive results obtained from standard two-dimensional (2D) *in vitro* genotoxicity assays (i.e., Ames test, mammalian cell gene mutation test, MNT and CA). Depending on the outcomes of the standard *in vitro* battery, which typically covers all genotoxicity endpoints (i.e., gene mutation, clastogenicity, and aneugenicity), a choice would be made between the RSMN assay or RS Comet assay (or both).

In recent decades, the use of animal testing for chemicals has come under increased scrutiny, prompting a global shift away from such studies due to concerns for animal welfare. Reflecting this, the European Union implemented the 7th Amendment to the Cosmetics Directive of the European Commission in 2003, prohibiting the use of *in vivo* follow-up testing for cosmetic ingredients since March 2009. In the context of a 'test battery,' this implies that a positive outcome from an *in vitro* standard genotoxicity assay would render the ingredient unsuitable for use in cosmetic products. Contemporary OECD genotoxicity TGs, such as OECD TGs 474 (OECD, 2016a), 488 (OECD, 2022), and 489 (OECD, 2016b), now underscore the importance of considering the intended or expected route of human exposure. Additionally, there is a growing emphasis on testing the 'site-of-contact,' as outlined in OECD TGs 488 (OECD, 2022) and 489 (OECD, 2016b). This is particularly relevant for ingredients in cosmetics, household products, pesticides, where the primary site of exposure is often the skin. All these points support the development and application of the RSMN and RS Comet assays.

(a) Analysis of the scientific rationale provided in the Test Submission

While the RSMN assay TST provides a cursory overview of the scientific rationale for test method development, the submission relies heavily on information previously published in scientific journals by the Test Submitters and the corresponding standard operating procedure (SOP) for the method. This format for a TST poses substantial challenges for reviewers because many information sources must be reviewed to properly assess the test method under consideration. Future submissions would benefit from including a summary of key information on scientific rationale, regulatory need, and

assay performance in the main TST document. Nonetheless, the submission package, which includes the TST and several informative publications, provide the necessary scientific rationale for the validation of this assay.

The MNT is designed to identify small, erratic (third) nuclei that are formed during the anaphase of mitosis or meiosis. Micronuclei are cytoplasmic bodies containing a portion of acentric chromosome or a whole chromosome which was not carried to the opposite poles during the anaphase. Micronuclei that remain after cell division are an indicator of permanent chromosomal damage. Micronuclei containing fragments of chromosomes indicate that the cell has undergone clastogenic DNA damage. Micronuclei containing whole chromosomes indicate that the cell has undergone aneugenic events. The MNT is used as a tool for genotoxicity assessment of various chemicals (Sommer et al., 2020).

Because the MNT requires cell division to occur, the assay can only be performed in dividing cells. *In vivo*, the MNT is usually performed on bone marrow. This comes with the advantage of a metabolically active model organism but requires analytical proof of internal exposure for the results to be toxicologically relevant. In contrast, traditional *in vitro* assays will always be subject to direct exposure but usually rely on rapidly dividing immortalized or cancer cell lines. These systems only feature limited metabolic competence which hence usually needs to be provided externally, for example by means of addition of or preincubation with S9-extract. The latter can be partially overcome by using more complex models, such as 3D human skin models. While not fully comparable, these models usually possess increased metabolic competence and do so with direct human relevance. However, initial efforts to transition to the full thickness skin models were unsuccessful since it contains mature keratinocytes, which are in a senescent state not suitable for micronucleus evaluation. Thus, the RhE model was ultimately proposed for use, as it still has proliferating cells, unlike full thickness skin models. However, it is important to note that use of this model with immature keratinocytes may limit the utility of this assay for testing metabolically activated chemicals, as immature keratinocytes do not express the full complement or capacity of metabolic enzymes.

The MNT has become a gold standard for rodent *in vivo* testing of chromosomal abnormalities (OECD TG 474 (OECD, 2016a)) and has long-standing regulatory recognition. The use of the MNT in conjunction with human-derived 3D skin models for *in vitro* genotoxicity testing therefore not only appears a straightforward choice for overcoming any species-specific effects but is also logical in terms of potential regulatory applicability as a component of tiered testing. Apart from the results of the validation trials, the TST draws support from its extensive literature references. Altogether this makes a solid case for the RSMN assay being a reasonable test for non-systemic dermal genotoxicity testing. Furthermore, the air-liquid-interface (ALI) culture conditions of the RSMN assay facilitates testing of a wide range of substances, including lipophilic substances at dose ranges relevant for topical exposure or application.

(b) Analysis of the regulatory rationale provided in the Test Submission

The regulatory rationale is described clearly in the TST. The need to develop reliable *in vitro* tests for genotoxicity testing associated with skin exposure was predominantly triggered by the European animal testing ban for cosmetics in conjunction with inherent challenges of alternative testing systems, one being the limited ability of existing cell-based *in vitro* methods to deliver sufficiently reliable predictions (e.g., elevated number of false positives). In addition, existing cell-based methods (i.e., MNT, CA, Ames test, and mammalian cell gene mutation tests) utilize rat liver S9 fraction to test for bioactivation of chemicals, which has an obvious limitation in the assessment of human skin genotoxicity. Citations provided in the TST (Pfuhler et al., 2020; Pfuhler et al., 2021) along with those identified by the ESAC (Bataillon et al., 2019; Wiegand et al., 2014) discuss the difference in metabolic enzyme expression in skin *vs.* liver and in rat *vs.* human, which provides a strong argument for the development a metabolically competent skin genotoxicity assay for assessment of human risk. These citations also provide evidence that the RSMN assay uses a metabolically competent human skin model.

The RSMN assay is intended to be used within regulatory genotoxicity hazard identification testing strategies to follow up positive results from the classical *in vitro* test battery, which is used as a first step. It is proposed to be used instead of *in vivo* genotoxicity test methods (i.e., OECD TG 474 (OECD, 2016a) and TG 475 (OECD, 2016c)) in the context of test items that are causing clastogenic (chromosomal breaks and translocations) or aneugenic (abnormal number of chromosomes) effects *in vitro* and are primarily associated with dermal exposure.

The TST makes a clear case for the performance of the test method and for the potential use of the RSMN assay as a standalone assay for detecting genotoxicity associated with skin exposure. However, as presented, the TST does not address concerns associated with potential future applications. For example, the ESAC sees the potential for using the combination of the RSMN and RS Comet assays as a substitute for *in vivo* genotoxicity assays and questions whether this use scenario will sufficiently address potential mutagenicity, or systemic effects from skin penetration and absorption. Importantly, the testing strategy is only being evaluated by the ESAC for the context of use, in which the RSMN assay is used in combination with the RS Comet for *in vitro* skin genotoxicity testing following topical exposure (see section 13).

1.3 Appraisal of the appropriateness of the study design

The TST compiles data from several published studies, thereafter, compiled to support the validation of the RSMN assay. This, together with the protracted timeline associated with completing the validation study, introduces challenges such as protocol modifications, change of participating laboratories, and varying numbers of technical and biological replicates. While a more targeted and concise approach would have been preferable, the overall design seems adequate with respect to critical parameters, including study design, laboratory independence, blinded sampling and independent result evaluation, all being sufficiently fulfilled. Moreover, the submitted SOP provides clear guidance on experimental design, result evaluation and substance calls.

1.4 Appropriateness of the statistical evaluation

The statistical methods used in the evaluation of the validation study data are standard, well-vetted methodologies and therefore appropriate for the purpose of evaluating the assay. However, no power calculation to support the number of chemicals necessary to assess the Within-Laboratory Reproducibility (WLR) or the Between-Laboratory Reproducibility (BLR) has been performed.

2. Collection of existing data

2.1 Existing data used as reference data

The *in vitro* test battery data and *in vivo* reference data were collected from the literature and the EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database (Kirkland et al., 2016; Corvi et al., 2018; Madia et al., 2020).

The RSMN assay data that are used for the validation/prediction model were produced during Phases 1 and 2 of the validation study. The results of Phase 1 were published by Aardema et al. (2010). The results of Phase 2 were published by Pfuhler et al. (2021). There are earlier publications which present data from the RSMN assay development (Curren et al., 2006; Mun et al., 2009) and that assess the transferability of the assay (Hu et al., 2009), but they were not used for the validation/prediction model. In addition, Kidd et al. (2021) conducted a study to validate the RSMN assay for use in their laboratory. No data from other sources were used in the validation report.

2.2 Existing data used as testing data

Not applicable. All data were the result of prospective testing.

2.3 Search strategy for retrieving existing data

Genotoxicity data were accessed from published literature and the EURL ECVAM database (Kirkland et al., 2016; Corvi et al., 2018; Madia et al., 2020). However, the search strategy was not specified in the TST.

2.4 Selection criteria applied to existing data

The selection criteria were based on the EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database.

3. Quality aspects relating to data generated during the study

3.1 Quality assurance systems used when generating the data

It was stated in subsection 3.5 of the TST that the acquisition of the data for the assessment of the reproducibility and predictive capacity did not follow any formal quality assurance system. Full Good Laboratory Practice (GLP) compliance is not a requirement for test method development or validation. The Test Submitters agreed to follow the safeguards recommended by Balls et al. (1995). The ESAC agrees with the quality assurance measures taken in this validation study.

3.2 Quality check of the generated data prior to analysis

As noted in the TST, acquisition of the data for the assessment of the reproducibility and predictive capacity did not follow any formal quality assurance system. However, several safeguards recommended by Balls et al. (1995) were applied, which the ESAC considers acceptable.

4. Quality of data used for the purpose of the study (existing and newly generated)

4.1 Overall quality of the evaluated testing data (newly generated or existing)

The quality of the data provided is high. Data are properly reported and appropriate quality criteria for analysing the data were provided. All underlying data used for analyses were provided with the TST.

4.2 Quality of the reference data for evaluating relevance[1]

The Test Submitters used chemicals supported by the literature and documented in the EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database.

4.3 Sufficiency of the evaluated data in view of the study objective

The strategy for test chemicals selection has not been clearly described or justified relative to the context of use described in the TST or the supporting documents. A balanced set of TP, TN, and MP would seem to be more appropriate for a standalone use of the method rather than confirmation of positives obtained using traditional genotoxicity assays as was the proposed context of use. Nonetheless, the ESAC considers the quality of the RSMN dataset sufficient to draw conclusions on assay validity (see subsections 4.1 and 4.2).

[1] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

5. Test definition (Module 1)

5.1 Quality and completeness of the overall test definition

Information provided regarding the biological and mechanistic relevance of the method is considered sufficient. When using the RSMN assay, test items are assessed in a commercially available 3D RS model (EpiDerm™ RhE model; MatTek; Ashland, MA, USA). The RSMN assay is biologically relevant for investigating DNA damage in the skin after topical exposure.

A detailed SOP was provided by the Test Submitters (see Attachment 1.1 provided with the TST) together with supporting published papers, clearly describing the test system used, the endpoint measured, the quality criteria applied to the test system, the acceptance criteria applied to the results, and the positive and negative controls. Over the years, several modifications of the protocol have been introduced to optimise it, including preparation and application of test items, aspects of experimental design, and other methodological considerations. The protocol recommended for future use foresees starting with a 72-h experiment, instead of a 48-h experiment to reduce the workload per substance, while maintaining the predictive capacity.

Genotoxicity potential is determined by measuring the frequency of micronucleated binucleated cells in the test item treated cultures compared to the solvent control for statistical significance and/or biological relevant increase (tissues with less than 60% (55±5%) survival are considered toxic and are not scored for micronuclei). To measure chromosomal damage, keratinocytes are isolated from the skin tissues and binucleated cells with micronuclei are scored, after acridine-orange staining using a fluorescence microscope. The percent binucleation is determined based on at least a 500-cell count. The validity criteria and assessment are clearly described in the SOP. Mitomycin C (3 µg/ml) is used as positive control. Negative controls include untreated tissues and a solvent control. Acetone or 70% ethanol (in water) are typically used as solvents.

5.2 Quality and completeness of the documentation concerning protocols and prediction models

The SOP of the RSMN assay evolved over several years and is supported by several publications. The SOP is well-described, including a description of the purpose, materials, study design, and assay procedure along with a description of how to terminate experiments and stain and score slides. Detailed data analysis instructions are also provided, including validity criteria.

6. Test materials

6.1 Sufficiency of the number of evaluated test items in view of the study objective

The main selection criterion for the initial selection of chemicals appears to be based on matched data from *in vivo* rodent skin exposure studies. This approach yielded a limited list of test chemicals that was later expanded to enlarge the substance pool for statistical and validation purposes, as well as to include more data rich substances. No specific criteria to select test items specifically for WLR were applied. Several Attachments (A2; A4, A5; A8, A10, A14) provided with the TST included information on mode of action and the list of test substances used in the different phases of the validation study. There were no statistical tests conducted at study initiation to justify the number of chemicals selected for the validation study. Nevertheless, the number of chemicals tested seems sufficient to assess the capacity of the method to detect expected modes of action and to evaluate reproducibility of the assay. The resulting list is fit for purpose in terms of the number of substances tested for assessing WLR and BLR. The ESAC noted that only a limited number of chemicals was selected for evaluating metabolic activation.

6.2 Representativeness of the test items with respect to applicability

The test items were selected for the validation study according to their description as TN, TP, and MP. The ESAC considers the test items sufficiently representative for the purpose of the study. Information was also provided on the genotoxic mode of action (e.g., alkylating agent). It should be noted that the TST does not make any statements related to defining the applicability domain or the representativeness of substance class/underlying mechanisms and coverage of technical parameters such as lipophilicity/solubility and stability.

7. Within-laboratory reproducibility (WLR) (Module 2)

7.1 Assessment of repeatability and reproducibility in the same laboratory

In the TST, it is stated that the validation study was not specifically designed for the WLR assessment, and for most cases only two experiments per test item are available, making it difficult to assess the WLR. The test items informing the WLR are identified in Attachment 4 of the TST.

No specific criteria to select test items specifically for WLR assessment were applied. The WLR was addressed by considering the test items for which at least two valid 48-h experiments and/or at least two valid 72-h experiments were available. As this was the case for 38 of the total 43 test items, they cover almost the entire validation set of the RSMN assay as explored so far. The number of test items is consistent with the number typically used in a validation study.

Results from two of the laboratories were consistent but only tested a small number of chemicals. Because the number of chemicals evaluated was so small, the ESAC also looked across laboratories and observed that results for Labs B and D were generally consistent while results for Lab A were inconsistent. Results produced by Lab C were generally consistent, but the number of test items evaluated was very limited and, for that reason, it is not possible to evaluate WLR for this laboratory. Based on data reported in Attachment 4, the ESAC re-calculated the reproducibility of results obtained at 48 h or 72 h separately (Tables 1 and 2).

Table 1. Reproducibility within each laboratory over time (WLR) in Phases 1 and 2 at 48 h

	Discordant	**Concordant**	**Total**	**%**
Lab A	6	13	19	**68%**
Lab B	2	18	20	**90**
Lab C	0	6	6	**100**
Lab D	1	13	14	**93**
All Labs	9	50	59	**85**
			Average	**88**

Note: Calculations are based on data obtained from at least two experiments.

The average of the WLR at 48 h based on data obtained from at least 2 experiments among the four laboratories is 88%.

Table 2. Reproducibility within each laboratory over time (WLR) in Phases 1 and 2 at 72 h

	Discordant	**Concordant**	**Total**	**%**
Lab A	1	3	4	**75**
Lab B	1	3	4	**75**
Lab C	3	0	3	**0**
Lab D	0	0	0	**-**
All Labs	5	6	11	**55**
			Average	**51**

Note: Calculations are based on data obtained from at least two experiments.
The average of the WLR at 72 h based on data obtained from at least 2 experiments among the four laboratories is 51%.

The tables below are based on data obtained from at least 3 experiments, which are required for the assessment of WLR, at 48 h and 72 h (Tables 3 and 4).

Table 3. Reproducibility within each laboratory at 48 h based on data from 3 experiments

	Discordant	Concordant	Total	%
Lab A	6	2	8	**25**
Lab B	1	1	2	**50**
Lab C	0	0	0	-
Lab D	1	4	5	**80**
All Labs	8	7	15	**47**
			Average	**51**

Table 4. Reproducibility within each laboratory at 72 h based on data from 3 experiments

	Discordant	Concordant	Total	%
Lab A	0	1	1	**100**
Lab B	1	0	1	**0**
Lab C	0	0	0	-
Lab D	0	0	0	**-**
All Labs	1	1	2	**50**
			Average	**50**

Based on three independent experiments for one test item (note the individual laboratories evaluated 1 – 8 test items), the average WLR for 48 h and 72 h is 51% (25-80%) and 50% (0-100%), respectively. Both of these values are well-below the 80% value considered acceptable for WLR. For the majority of test items, only two experiments are available. To better understand the robustness of results for the same test item across laboratories, the concordance of experiments for the same test items tested in the different laboratories was calculated for both 48 h and 72 h (Tables 5 and 6). Based on these calculations, concordance for 48 h and 72 h is 72% and 33%, respectively.

Table 5. Concordance of experiments for the same chemicals tested in the different laboratories at 48 h

	Discordant	Concordant	Total	%
All Labs	5	13	18	**72**

Table 6. Concordance of experiments for the same chemicals tested in the different laboratories at 72 h

	Discordant	Concordant	Total	%
All Labs	2	1	3	**33**

7.2 Conclusion on within-laboratory reproducibility as assessed by the study

Typically, three independent experiments are required for the assessment of WLR. For the RSMN assay, most test items used by the Test Submitters to assess WLR have only been evaluated in two experiments. The ESAC analysed the data for those test items that had three or more experiments at either 48 h (Table 3) or 72 h (Table 4). For Lab A in which 8 test items were evaluated at 48 h, 6 are discordant. In Lab B, 1 of 2 test items are discordant at 48 h. No test items were evaluated in Lab C. Finally, in Lab D, 1 of 5 test items is discordant at 48 h. Based on three independent experiments for one test item, the average WLR for 48 h and 72 h is 51% (25-80%) and 50% (0-100%), respectively. Both values are well-below the 80% value considered acceptable for WLR. When calculated based on three independent experiments, concordance is very poor, especially in Lab A. Concordance of experiments for the same test items evaluated in the different laboratories at 48 h is 72% (Table 5), which is still below the target of 80% for concordance. Data at 72 h are insufficient to draw any conclusions on the WLR. Consequently, there is a high degree of uncertainty for WLR calculations based on the results of only two experiments because there is a high probability that a third experiment would be discordant. To better understand these results, the ESAC sought input from the Test Submitters. The Test Submitters stated that there are no obvious reasons for most of these discordances but described several factors related to standardisation of the protocol that may impact WLR (see TST).

While the WLR based on data obtained from at least two experiments among the four laboratories is acceptable (68-100%), the WLR based on three independent experiments is not (25-80%). Thus, based on the available data, ESAC cannot conclude on the WLR. However, it is important to remember that the RSMN assay will be used in a tiered approach and in practice the yes/no classification will be based on the majority of calls. For example, two positive experiments will be sufficient to classify the test item as positive for genotoxicity, the results of the third experiment will not change the classification. For that reason, the WLR of the RSMN is acceptable. In addition, a recent study (Thakkar et al., 2022) evaluating 22 fragrance materials in the RSMN assay provided valuable insight and enhanced confidence in the assay. All things considered, the proposed strategy, consisting of an initial assay utilizing a 2-day dosing regimen (48-h treatment), followed by a confirmatory assay using a 3-day dosing regimen (72-h treatment) if the initial result is negative, is deemed appropriate.

8. Transferability (Module 3)

8.1 Quality of design and analysis of the transfer phase

Successful transfer of the RSMN assay was demonstrated in the publications by Hu et al. (2009) and Aardema et al. (2010). The design and organization of these demonstrations are summarized below.

Preliminary transfer - Early in the development of the RSMN assay, it was successfully transferred using a protocol similar to the SOP used during the validation study from the developing laboratory (i.e., Institute for *In Vitro* Sciences, Inc. (IIVS)) to two American laboratories (Procter and Gamble and MatTek). Five test items, three positives and two negatives, were correctly predicted by all three laboratories during transfer. Two additional negative test items were tested by only two of the laboratories and were also correctly predicted (Hu et al., 2009). See Attachment 5 of the TST for the specific test items used.

Validation study - The RSMN was formerly transferred to naïve labs as part of Phase 1 of the validation study using a protocol detailed in Dahl et al. (2011). This protocol was “basically the same as used in Phase 2 of the validation study. However, relative viable cell count was not yet used as a cytotoxicity readout and a 72-h time point was not used to follow up results of the 48-h time point.” Using this protocol, the RSMN assay was also successfully transferred to two naïve European laboratories (i.e., L’Oréal and Henkel). Successful transfer was demonstrated using over-sea shipped tissues testing the two positive test items vinblastine sulfate and MMC. The results have been reported by Aardema et al. (2010).

Establishment of the SOP for conducting the RSMN assay - The Test Submitters provided an SOP (Attachment 1 of the TST) to follow when using the RSMN assay moving forward. The SOP clearly describes the test system used, the endpoint measured, the quality criteria applied to the test system, the acceptance criteria applied to the results, and the positive and negative controls. This SOP was used in the later phases of the validation study, but not in the transferability phase.

The ESAC wants to draw attention to the following points: to conduct the RSMN assay, laboratories must possess standard technical laboratory skills and the ability to follow OECD TGs as well as expertise in interpreting data and results of genotoxicity tests. Experience with 3D tissue handling and with defining the dose range for *in vitro* genotoxicity tests is particularly important.

As noted in the TST, successful training should be demonstrated by testing at least one positive, or preferably one positive and one negative item, following the entire protocol from determination of solubility to the final call. Test items should be selected from the list of items tested during the validation exercise. Note that additional experiments are required to establish a reliable laboratory-specific historical control range.

8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study

Transferability studies are limited and were conducted using protocols that were ultimately converted into an SOP; however, they are sufficient to demonstrate transferability of the RSMN assay to naïve laboratories.

9. Between-laboratory reproducibility (BLR) (Module 4)

9.1 Assessment of reproducibility in different laboratories

The test items informing BLR are identified in Attachment 8 of the TST. BLR was addressed in both phases of the validation study. In Phase I, chemicals were selected by an independent chemicals selection expert team (Aardema et al., 2010). In Phase 2, they were initially selected by external experts (i.e., R. Corvi and D. Kirkland) from a chemical master list prepared for Cosmetics Europe. In total, 22 test items informed the BLR. Chemicals were selected to obtain a balanced selection with respect to the test items reference result (11 TP, 4 MP, and 7 TN) and to address various modes of action. The substances were mainly organic representing various chemical classes and many functional groups. In addition, two metal salts were included.

Three of the 22 test items were also used during the development of the RSMN assay (MMC by Curren et al., 2006; n-ethyl-n-nitrosourea and 4-nitrophenol by Mun et al., 2009). In addition, some test items were also used for the assessment of the transferability (MMC by Hu et al., 2006 and Aardema et al., 2010; methyl methanesulfonate and 4-nitrophenol by Hu et al., 2009).

Sixteen test items had concordant calls and 4 test items (cadmium chloride, 2,4-diaminotoluene, 5-fluorouracil, phenantrene) had discordant calls in two to four laboratories. The remaining two test items (i.e., resorcinol, tolbutamide) were negative in one or two laboratories and equivocal in one laboratory.

The overall BLR calculated by the Test Submitters was **16/22 = 73%**. The ESAC noted that this is slightly below the 80%, the value considered acceptable for BLR.

The Test Submitters identified experiments with borderline results and differences in dose selection as well as cytotoxicity between laboratories as potential experimental sources of variability that can result in reduced BLR.

9.2 Conclusion on between-laboratory reproducibility as assessed by the study

Typically, three independent laboratories are required for the assessment of BLR. For the RSMN assay, 15 test items had a call (i.e., negative or positive result) in at least three laboratories for the 48-h time point. The remaining seven test items had a call from only two laboratories. Only one test item had data for the 72-h time point in three laboratories. Therefore, the ESAC analysed the data for test items that gave results in three or more laboratories for only the 48-h time point. This results in 15 test items of which 11 were reproducible between laboratories and 4 were not, resulting in a similar overall BLR as reported by the Test Submitters.

10. Predictive capacity and overall relevance (Module 5)

10.1 Adequacy of the assessment of the predictive capacity in view of the purpose

Overall, a limited number of chemicals (i.e., 43 test items) was used to assess the predictive capacity of the RSMN assay (Attachment 9 provided with the TST). Information on mode of action, reference *in vitro* and *in vivo* genotoxicity data, as well as carcinogenicity data are reported in Attachment 10 of the TST. For two TP test items listed in Attachment 10, there is no indication of *in vivo* genotoxicity data (i.e., cyclopenta[c,d]pyrene and 4-vinyl-1-cyclohexene diepoxide). The chemicals were, with two exceptions (i.e., cadmium chloride and potassium bromate), organics representing various chemical classes and featuring numerous functional groups.

Test items were selected to cover a variety of modes of action and to include three categories: TN (10 test items) and TP (21 test items), with concordant *in vitro* and *in vivo* data, and MP (12 test items) for which positive *in vitro* findings were reported but not confirmed in *in vivo* studies. Two of the test items used for the optimisation of the test method were also used in the assessment of predictivity. Of the 21 TP test items, 7 needed metabolic activation to show positive results (Attachment 10 provided with the TST).

The predictive capacity of the RSMN assay was analysed using two different approaches and the related results were reported in the TST as well as in Attachment 9 provided by the Test Submitters. Three out of seven TP that require metabolic activation were tested in more than one laboratory and allowed for between laboratory concordance assessment (i.e., 2-AAF, 2,4-DAT, and 4-vinyl-1-cyclohexene diepoxide). Only one test item (i.e., 4-vinyl-1-cyclohexene) showed a positive study outcome in all three laboratories. Two metabolically active test items (i.e., CPPE and diethylstilbestrol) were only tested in one laboratory and showed a positive study outcome. Another two pro-mutagens (i.e., 2-AAF and DMBA) were not identified as TP (tested in three and one lab(s), respectively). These results indicate that metabolic activation may not be sufficient in the RS model to form micronuclei. Other false negative test items or those not consistently classified among the different laboratories were cytosine arabinoside, cadmium chloride, 5-fluorouracil. Among the MP, curcumin and eugenol were classified as positive (false positive), while all the other MP were correctly classified as negative. Among the TN, diclofenac and phenanthrene were classified as positive (false positive).

False predictions have been identified and extensively discussed in paragraph 2.5.3 of the TST. Thirteen test items were not correctly predicted in at least one laboratory. In about half of the cases, a possible explanation for misclassification has been suggested by the Test Submitters.

In a publication made later available by the Test Submitters (Thakkar et al., 2022), additional MP test items were tested, and all were properly classified as negative by the RSMN, further supporting the high specificity of the assay. However, the sensitivity of the RSMN in some of the participating laboratories (i.e., Lab B and Lab D) was very low: 8/13 (61.5%) for Lab B and 2/4 (50%) in Lab D, which is of concern as this indicates a possible high rate of false negative results. This together with the low predictive capacity toward chemicals requiring metabolic activation points out to some limitations of the assay that requires caution in the classification. Among the TP test items with a univocal positive/negative classification, the sensitivity over test items is 14/21 (67%). Considering the risk associated with incorrect classification, before concluding on the negativity of a test item, the ESAC recommends using the RSMN in combination with the RS Comet assay and together with any other additional information before concluding on the negativity.

10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

Even if generated by using a limited number of chemicals, conclusions on the use of the RSMN assay as a 'surrogate' of the usual *in vivo* follow-up for *in vitro* positive test items are supported by the available data, with some limitations. The incorrect predictions need be viewed in the context of the proposed application of the RSMN assay, i.e., as one of three higher tier test methods (RS Comet assay and Hen's Egg Test for Micronucleus Induction) to follow-up potentially incorrect positive results from the standard *in vitro* test battery.

Overall, the false negative chemicals were 33.3% (7 out of 21), and the false positive 27.3% (6 out of 22), with a clear improvement concerning the specificity compared to the standard *in vitro* MNT, which has a specificity of approximately 30% (Kirkland et al., 2005). In addition, among the 12 MP test items included (positive *in vitro* and negative *in vivo*), 10 were correctly classified as negative and only two were judged positive (i.e., curcumin and eugenol). This provides an acceptable rational of the use of the RSMN assay as a follow up from the standard *in vitro* test battery positive results as a surrogate of *in vivo* methods for dermal applied compounds. The latest being particularly relevant for those regions in which cosmetic ingredients cannot be tested *in vivo*, e.g., the EU. However, the sensitivity of 67% (14/21) requires caution in the interpretation as TP maybe wrongly classified as negative.

While the 72 h has been used to increase the sensitivity, it is not clear how the specificity would be affected because only a few TN and MP test items have been tested at 72 h. However, in a recent publication (Thakkar et al., 2022), the added value of the 72-h time point has been clearly demonstrated. In this paper, a set of 22 fragrance materials were evaluated in the RSMN assay. For these materials, *in vitro* as well as *in vivo* MNT data are also available. The results of the RSMN assay were in 100% agreement with results from the *in vivo* MNT. In the dose range-finding assay and the first main study, tissues were treated with a 2-day dosing regimen (48-h harvest). In the confirmatory assay, tissues were treated with a 3-day dosing regimen (72-h harvest). Each chemical was tested in triplicate, using tissues generated from the same batch/skin donor and results from both 48 h and 72 h are shown. Thus, these results are useful and increase the confidence on the RSMN assay. The strategy used (the current study design for the RSMN assay involves an initial assay using 2-day dosing regimen, and if the result is negative then a follow-up confirmatory assay using 3-day dosing regimen is conducted) is proper.

The Test Submitters stated that in the bridging study one third of all 48-h TN and MP test items (6/18) were tested at 72 h and all six were negative. The Test Submitters therefore do not expect a different response between the TN at 48 h and 72 h. The ESAC can only partially concur with this statement as only 4 test items were negative in all experiments performed in a laboratory.

Another limitation identified is the predictive capacity toward test items requiring metabolic activation, as only one out of seven TP metabolic active test item could consistently be identified in more than one laboratory.

Overall, the predictive capacity and overall relevance of the chemicals tested at 48 h, with exception of the ones that need metabolic activation, is adequate even if not all laboratories were performing equally well. For the chemicals tested at 72 h, conclusions cannot be made due to the limited amount of data available.

11. Applicability domain (Module 6)

11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

For organizational purposes, the ESAC discusses the strengths and limitations of three different aspects of the applicability domain below.

Biological domain: As proposed in the TST, the RSMN assay would be used for screening chemicals for potential genotoxicity resulting from dermal application of the chemical. The test was designed primarily for the purpose of screening cosmetic ingredients and is therefore focused on the relevant portal of entry (i.e., skin). Use of 3D RS allows for a more biologically relevant tissue architecture, which is more likely to recapitulate *in vivo* chemical penetration and toxicological response. Based on the evidence provided in the TST, the ESAC concurs that the assay is useful for the purpose of identifying chemicals with the potential to cause genotoxicity following dermal application – in a tiered testing paradigm that includes the standard *in vitro* genotoxicity assays and the RS Comet assay as described in section 1. However, for the broader purpose of human health risk assessment, additional evidence would be required to make a regulatory decision, as this assay lacks the ability to account for potential systemic effects from dermal exposure.

Chemical domain: The applicability domain of the test items was not specifically explored in the TST. However, there are clear advantages to the use of a 3D RS model that is cultured at the ALI compared to traditional genotoxicity assays performed in submerged 2D monocultures, particularly in terms of the potential to increase the chemical applicability domain. In the case of the RSMN assay, the use of EpiDerm™ RhE tissues also has the potential to be useful for identifying metabolically activated chemicals in a human relevant context. This attribute is a significant advantage over immortalized or cancer cell line-based assays currently used in the traditional 2D submerged assays that rely on rat liver S9 fraction to approximate human skin metabolism.

Evidence of metabolic competence and ability of the RSMN assay to identify bioactivated genotoxicants is provided in the TST and referenced publications. Out of 15 test items used as TP chemicals, 4 bioactivated genotoxicants were evaluated during the validation study. In addition, because the model involves culturing at the ALI, a broader universe of chemicals and substances that are not generally well-tolerated by submerged *in vitro* assays can be tested (e.g., lipophilic and otherwise poorly dissolved in cell culture medium). Furthermore, complex mixtures and formulations in various vehicles (creams, emulsions, etc.) could potentially be tested in a skin model cultured at the ALI. Since the TST did not specifically evaluate lipophilic chemicals or product formulations, no conclusions can be drawn on the utility of this assay for such purposes. However, the design of the assay suggests the potential to expand beyond the limited chemical space evaluated in the validation study, which defines all submerged 2D *in vitro* systems.

Mechanistic domain: Regarding the chemical space covered by the chemicals evaluated in the validation study, the TST would have benefitted from additional details regarding underlying mechanisms, substance class representation, and test chemical stability. Hence, there remain some uncertainties regarding the reasoning of some of the substance selections. A recently developed AOP for genotoxicity (Sasaki et al., 2020) provides a streamlined, science-backed series of key events for various mechanisms of chemical-induced genotoxicity and would be a valuable resource for better defining the domain of applicability for this assay in terms of mode of action.

Unlike the RS Comet, the RSMN assay is not expected to have fully mature keratinocytes because it utilizes the EpiDerm™ RhE model. As such, it is unlikely that the RSMN possesses maximal metabolic competence. To that point, most of the pro-mutagens (e.g., DMBA, 2-AAF, CPPE, 2,4-DAT) included in the validation study to explore the potential of the RSMN assay to detect substances that need to be metabolically activated to show their mutagenic potential tested negative in at least one laboratory. Thus, caution must be used when interpreting negative results obtained with test items that require metabolic activation.

Overall, the predictive capacity, WLR, and BLR of the test items evaluated at 48 h is adequate, except for test items that require metabolic activation. In addition, of the seven TP test items tested to evaluate the ability of the RSMN test to identify bioactivated genotoxicants, only three were tested in more than one laboratory and only one test item showed a positive response in all three test laboratories. Based on the data provided in the TST, the ESAC concluded that the metabolic competence of the RSMN may not be sufficient, possibly due to use of the EpiDerm™ RhE model that utilizes immature keratinocytes. While a number of peer-reviewed publications provided with the TST showed evidence for the existence of native metabolizing enzymes in the EpiDerm™ RhE tissues, the ESAC cannot draw conclusions on the utility of the RSMN alone for testing potentially bioactivated genotoxins with the data in-hand.

However, all seven bioactive TP test items, except one, that were tested in both RS Comet and RSMN were correctly classified in at least one assay. The only outlier (i.e., diaminotoluene) is still classified correctly with a strategic combination score of 0.33 instead of 0. These results support the use of the tiered testing paradigm proposed in the TST because it is possible to correctly identify test items that require bioactivation when using the RSMN and RS Comet assays in combination. Therefore, the ESAC concludes that the combination of RSMN and RS Comet is preferred for general use and for substances that require bioactivation.

11.2 Quality of the description of applicability domain, limitations, exclusions

This study did not attempt to define the applicability domain in terms of chemical space. As such, this cannot be evaluated by the ESAC.

12. Performance standards (Module 7)

12.1 Adequacy of the proposed Essential Test Method Components

Not applicable.

12.2 Adequacy of the proposed Reference Chemicals

Not applicable.

12.3 Adequacy of the proposed performance target values

Not applicable.

13. Readiness for standardised use

13.1 Assessment of the readiness <u>for regulatory purposes</u>

The RSMN assay should be considered together with the RS Comet assay, as they are intended to be used within tiered regulatory genotoxicity hazard identification testing strategies to follow up positive results from the classical *in vitro* test battery. The RSMN assay is proposed to be used instead of *in vivo* genotoxicity test methods (i.e., OECD TG 474 (OECD, 2016a) and TG 475 (OECD, 2016c)), in the context of test items that are causing clastogenic (chromosomal breaks and translocations) or aneugenic (abnormal number of chromosomes) effects *in vitro* and are primarily associated with the dermal exposure route. In combination with the RS Comet assay, they cover all genotoxicity endpoints that usually need to be addressed for regulatory purposes (i.e., gene mutation, clastogenicity, and aneugenicity).

The approach proposed by the Test Submitters involves conducting the RSMN assay as a follow-up to a positive *in vitro* MNT or CA, and a negative result in the RSMN assay would support overriding the initial positive result. Similarly, a positive Ames and/or mammalian cell gene mutation tests would trigger a follow-up with the RS Comet assay, and a negative result in the RS Comet assay would support overriding the initial positive result. According to the Test Submitters, in cases where both Ames/mammalian cell gene mutation tests and MNT/CA studies yield positive results, follow-up testing should be conducted with both the RS Comet and RSMN assays, and a positive follow-up result in either the RS Comet assay or the RSMN assay would lead to the conclusion of a genotoxicity hazard associated with topical exposure.

As a precautionary note, test items requiring metabolic activation may be classified as false negatives based on RS genotoxicity tests alone. This is particularly crucial when these compounds are accessible for systemic absorption. Considering the skin's lower metabolic capacity, initial results might appear negative. However, once absorbed, chemicals could undergo systemic bioactivation, thereby elevating the risk of neoplasia, which is an unacceptable outcome.

The 12th Revision of the Scientific Committee on Consumer Safety (SCCS) Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation, concluded that (Pfuhler et al., 2020; Pfuhler et al., 2021; SCCS, 2023):

a. 3D tissue models simulate *in vivo*-like conditions including cell viability, proliferation, differentiation, morphology, gene and protein expression. These models can complement classical 2D cell culture-based assays;
b. 3D tissue-based genotoxicity assays can be used as 2nd tier assays to follow-up on positive results from standard *in vitro* assays;
c. For adoption of a tissue model as a 2nd tier assay, ability to detect the full range of genotoxic damage (leading to mutagenicity, clastogenicity, aneugenicity) should be demonstrated;
d. The 72-h protocol for the RSMN has higher sensitivity than the 48-h protocol;
e. The RS Comet assay (Pfuhler et al., 2020) and RSMN (Pfuhler et al., 2021) assays are now sufficiently validated to move towards individual OECD TGs, but an independent peer review of the validation study is still needed.

13.2 Assessment of the readiness <u>for other uses</u>

For the intended purpose, no mechanistic limitations of the RSMN assay are known, possibly except for metabolic activation. Since only a very limited number of chemicals have been tested (n = 43), its applicability for screening purposes will need to be evaluated further. The applicability to mixtures and final formulations has not been demonstrated yet.

13.3 Critical aspects impacting on standardised use

In the validation study, which was conducted over a protracted period of time, the composition of the participating laboratories changed. Still, the method has been optimized and offers several advantages compared to traditional cell cultures. No critical aspects impacting standardized use have been identified. The only technical limitation identified relates to test item solubility. The suitability of solvents other than acetone and 70% ethanol, has not been tested. Finally, the Test Submitters recommend only using the 72-h time point. More experiments should be performed at the 72-h time point for the ESAC to be fully confident that 72 h is the only time point required.

13.4 Gap analysis

Even if different chemical classes have been tested, due to a small number of test items (n = 43), the applicability of RSMN assay will need to be evaluated further. Test items that require metabolic activation may give false negative results.

14. Other considerations

The ESAC sees value in potential expanded use for this method and the combination with the RS Comet for the evaluation of other product categories and configurations (e.g., pesticides, chemical mixtures, final products), provided systemic bioavailability is considered.

The RSMN and RS Comet assays were designed for the follow up of a positive MNT and/or CA, and/or a positive Ames and/or mammalian cell gene mutation tests performed as part of the traditional *in vitro* genotoxicity test battery, respectively. As such, the assays can aid with weight-of-evidence (WoE)-based regulatory decision-making for potentially genotoxic substances with skin contact. In this context, the use of 3D RS models not only offers the advantage of organ similarity with regard to barrier function but also (limited) metabolic competence. Unlike other test systems, however, the models also provide human relevance. It should be noted that, as presented by the Test Submitters, the system will not allow any predictions on potential systemic genotoxicity. Accounting for systemic genotoxicity will require additional considerations such as the use of liver S9 fraction and/or data on skin absorption/penetration. In cases of significant skin absorption of a compound found to be positive in traditional genotoxicity assays following metabolic activation (by S9) but negative in the RS assays, the possibility of its liver activation into a genotoxic compound cannot be excluded due to potential differences in the metabolic capacity of the skin and the liver. Hazard classification should be straight forward for all substances identified as initial positives which are not subject to S9-activation as these will either be confirmed or overruled by subsequent testing with RS assays. However, for substances originally requiring S9-activation, overruling by the RS assays is less straight forward as a negative result obtained using the RS assays cannot principally exclude the possibility of systemic activation following skin penetration. Therefore, such substances will have to undergo further testing followed by WoE analysis.

15. Conclusions on the study

15.1 ESAC SG summary of the results and conclusions of the study

General comments

The ESAC agrees with the proposed use of the RSMN assay as a confirmatory assay to follow up positive *in vitro* MNT and/or CA findings following dermal application for potential genotoxic chemicals.

The RSMN assay is not currently proposed as a stand-alone assay for systemic genotoxicity. If it were to be considered for use in systemic genotoxicity assessment, data derived from the assay would need to be considered together with other data in a WoE approach.

The ESAC had numerous discussions on the role of this assay in a WoE approach. A demonstration of some of the considerations that the ESAC believes are important are discussed in section 14 above. These considerations include, but are not limited to, the degree of metabolic activation and potential systemic availability.

The ESAC sees value in potential expanded use for this method and the combination with the RS Comet assay. See other considerations (section 14) for additional discussion.

Chemicals selection

The strategy for chemicals selection has not been clearly described or justified relative to the context of use described in the TST or in the supporting documents. To this point, there appears to be a disconnect between the proposed purpose of the assay (i.e., confirmation of positive results obtained in the standard *in vitro* genotoxicity battery) and the chemicals selection strategy. A balanced set of TP, TN, and MP substances would seem to be more appropriate for a standalone use of the methods rather than the proposed use of confirming positive results. The ESAC also notes that the test items were selected and classified as TP, TN, and MP by subject matter experts that were not part of the validation study and that the rational for those decisions was not provided in the TS. Consequently, there is some uncertainty associated with these classifications that the ESAC could not resolve.

Validation study design

The TST compiles data from several published studies, thereafter, compiled to support the validation of the RSMN assay. This, together with the protracted timeline for completing the validation study, introduce challenges such as protocol modifications, including a change in the duration of the assay (i.e., 48-h and 72-h time points), and differences in numbers of technical and biological replicates that were tested in the participating laboratories. While a more targeted and concise approach to validation would have been preferable, the overall design seems adequate with respect to critical parameters such as study design, laboratory independence, blinded sampling, and independent result evaluation. Moreover, the submitted SOP provides clear guidance on experimental design, result evaluation and substance calls.

48-h vs. 72-h time points

Overall, the predictive capacity and WLR and BLR of the test items evaluated at 48 h is adequate, with exception of the ones that need metabolic activation. For the chemicals tested at 72 h, conclusions cannot be made due to the limited amount of data available in the TST. The Test Submitters suggested there may be an increased sensitivity of the RSMN assay at the 72-h time point compared to the 48-h time point. In support of this suggestion, after the submission of the validation study to EURL ECVAM, a paper was published (Thakkar et al., 2022) that clearly demonstrates the significance of the 72-h time point. This paper was provided to the ESAC by the Test Submitters to support the peer review of the RSMN assay. In this study, 22 fragrance materials with *in vitro* and *in vivo* MNT data underwent evaluation in the RSMN assay. Initially, tissues were treated with a 2-day

dosing regimen (48-h harvest) in the dose range-finding and primary studies. Subsequently, a confirmatory MNT utilized a 3-day dosing regimen (72-h harvest). Each chemical underwent triplicate testing using tissues from the same batch/skin donor, with results presented for both 48 h and 72 h. The RSMN assay results exhibited complete concordance (100%) with the *in vivo* MNT findings with the 48-h time point followed by the 72-h confirmatory time point. These findings enhance the assay's utility and reliability, affirming the appropriateness of the study's design involving a sequential dosing regimen approach.

Metabolic competence of RSMN assay

The literature supports the idea that RS models should be useful for evaluating metabolically active compounds. However, based on the data provided by the Test Submitters, we cannot draw a conclusion on this aspect of the RSMN assay's utility. The RhE model used in the RSMN assay is not expected to have fully mature keratinocytes. As such, it is unlikely that model has maximal metabolic competence. In the TST, 2 chemicals (i.e., 2-AAF, DMBA) that tested positive with the addition of rat S9 to the *in vitro* genotoxicity assays were negative in the RSMN assay. A third pro-mutagen, 2,4-diaminotoluene, was inconclusive in RSMN assay. Evaluation of the use of the RSMN assay for identifying metabolic activation is hampered by the limited number of chemicals tested in more than one laboratory which required metabolic activation, as well as the potential differences between traditional methods that use rat liver S9 fraction, compared to the RSMN assay that attempts to account for native skin metabolizing enzymes. Therefore, unless there is additional complementary data to support interpretation of results, caution is necessary when using the RSMN assay for test items that require metabolic activation. However, it should be noted that since these compounds tested positive in the *in vitro* Ames as well as other assays (MNT, CA test), the proposed testing scheme would also require these test items be evaluated in the RS Comet assay.

Based on the available data, neither the RSMN assay (metabolic competence may be limited due to maturation state of keratinocytes) nor the RS Comet assay (not enough data to conclude on potential to correctly classify metabolic active test items) demonstrated success identifying metabolically activated genotoxicants. However, the combination of the RS Comet and RSMN assays correctly identified six out of seven TP test items that require bioactivation (strategic combination of 1). The only outlier (i.e., diaminotoluene) was negative in the RS Comet and positive in 1 out of 3 laboratories in the RSMN (strategic combination of 0.33). Therefore, also for bioactive chemicals, using the combination of the RSMN and RS Comet assays is preferred.

WLR and BLR

The validation study was not designed to evaluate WLR or BLR. The number of test items evaluated in three or more independent experiments in each laboratory was very limited and, in some cases, test items were only evaluated in one laboratory. Focusing on the WLR, the concordance based on test items where three experiments were performed in an individual laboratory is very poor, especially in tests performed by Lab A. To increase the number of comparisons for each chemical, the ESAC considered replicates across the laboratories (excluding Lab A). With this approach, ESAC was able to evaluate the concordance of experiments for the same test items evaluated in different labs at the 48-h time point. When applying this approach, the agreement for a test item evaluated across replicates increased from 54% to 72%, which is still borderline because WLR is generally expected to be at least 80%. While that is the case, this value is more acceptable than the previous value of 54% that was based on individual laboratory results at the 48-h time point. The overall BLR calculated by the Test Submitters was 73%. ESAC noted that this is slightly below 80%, the value considered acceptable for BLR.

15.2 Extent to which study conclusions are justified by the study results alone

Data presented in this study were published in peer reviewed publications. The RSMN assay was designed primarily for the purpose of screening cosmetic ingredients and is therefore focused on portal of entry, specifically skin. Based on the evidence provided in the TST, the ESAC concurs that the assay is useful for the purpose of identifying chemicals with the potential for genotoxic effects following dermal exposure when used in a tiered testing paradigm that includes the standard *in vitro* genotoxicity test battery and the RS Comet assay as described in section 1. However, for the broader purpose of human health risk assessment, additional evidence would be required to make a regulatory decision, as this assay lacks the ability to account for potential systemic effects from dermal exposure.

15.3 Extent to which conclusions are plausible in the context of existing information

All published data generated using the RSMN assay have been included in the TST. Additionally, the ESAC is aware that the RSMN assay has been used to generate data used to support the regulatory assessment of cosmetic ingredients.

16. Recommendations

16.1 General recommendations

The validation study would have benefited from applying a more comprehensive, a priori validation strategy. The ESAC sees validation as a process of continuous confidence building rather than a single ring trial study. To that point, the ESAC does not see a problem with how this study was conducted (i.e., multiple steps). It is, however, critical that studies like the one evaluated here are coherent across study phases to promote confidence in the data. Specific recommendations related to the validation and application of the RSMN assay include:

- Apply a more mechanistically driven approach for chemicals selection. To ensure sufficient coverage of the chemical space for the intended purpose of the method, it would have been helpful if the chemicals selected for the validation were coordinated at study initiation to evaluate, for example, metabolism and *in vitro* biokinetics.
- To assess transferability to naïve laboratories the number of test items, and independent experiments that need to be performed needs to be carefully considered.
- Predictive capacity was determined using a relatively small set of chemicals (i.e., 43). Including more chemicals in the validation would have been preferred.
- As a best practice, three replicate experiments within a laboratory should be used for a validation study to assess WLR.
- As a best practice, at least three laboratories should be used for a validation study to assess BLR.
- Due to the limited number of test items evaluated, additional studies are needed to fully define the applicability domain. To that point, the ESAC recommends further evaluating the utility of the RSMN assay, including more compounds that require bioactivation. This could be accomplished by using more publicly available data as it becomes available.
- The suitability of solvents for use in the RSMN assay other than acetone and 70% ethanol should be investigated and guidance for selection of solvents for specific test items should be included in the SOP.
- Future applications:
    - The assay is fit for the application described in the TST. Because the method provides valuable information for compounds applied topically to the skin in general, the use application could potentially be broadened and the ESAC would recommend validation for other topically applied product classes such as mixtures, and formulations.
    - Due to differences in skin and liver metabolism, the ESAC expresses some concern related to the possibility that a topically applied chemical that penetrates the skin may become systemically available and could be bioactivated by the liver. End users cannot always assume that a compound applied topically to the skin will not be systemically available and metabolized by the liver. Hazard classification should be straight forward for all substances identified as initial positives which are not subject to S9-activation as these positive results will either be confirmed or overruled by subsequent RS-testing. However, for substances originally requiring S9-activation overruling by the RS assays is less straight forward as a negative RS-result cannot principally exclude the possibility of systemic activation following skin penetration. Therefore, such substances will have to undergo further testing followed by WoE analysis (see other considerations, section 14).

16.2 Specific recommendations (e.g., concerning improvement of protocols)

Not applicable.

17. References

- Aardema MJ, Barnett BC, Khambatta Z, Reisinger K, Ouedraogo-Arras G, Faquet B, Ginestet AC, Mun GC, Dahl EL, Hewitt NJ, Corvi R, Curren RD (2010) International prevalidation studies of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. Mutat Res 701(2):123-31. doi: 10.1016/j.mrgentox.2010.05.017. https://www.sciencedirect.com/science/article/pii/S1383571810002007?via%3Dihub.

- Balls M, Blaauboer BJ, Fentem J, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, Reinhardt CA, Repetto G, Sladowski D, Spielmann H, Zucco F (1995) Practical aspects of the validation of toxicity test procedures: the report and recommendations of ECVAM workshop 5. Altern Lab Anim 23(1):129-147. doi: 10.1177/026119299502300116. https://journals.sagepub.com/doi/10.1177/026119299502300116.

- Bataillon M, Lelièvre D, Chapuis A, Thillou F, Autourde JB, Durand S, Boyera N, Rigaudeau AS, Besné I, Pellevoisin C (2019) Characterization of a new reconstructed full thickness skin model, T-Skin™, and its application for investigations of anti-aging compounds. Int J Mol Sci 20(9):2240. doi: 10.3390/ijms20092240. https://www.mdpi.com/1422-0067/20/9/2240.

- Corvi R, Madia F (2018) EURL ECVAM genotoxicity and carcinogenicity consolidated database of Ames positive chemicals. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames.

- Curren RD, Mun GC, Gibson DP, Aardema MJ (2006) Development of a method for assessing micronucleus induction in a 3D human skin model (EpiDerm). Mutat Res 607(2):192-204. doi: 10.1016/j.mrgentox.2006.04.016. https://www.sciencedirect.com/science/article/pii/S1383571806001008?via%3Dihub.

- Dahl EL, Curren R, Barnett BC, et al. (2011) The reconstructed skin micronucleus assay (RSMN) in EpiDerm™: detailed protocol and harmonized scoring atlas. Mutat Res 720:42-52.

- Hu T, Kaluzhny Y, Mun GC, Barnett B, Karetsky V, Wilt N, Klausner M, Curren RD, Aardema MJ (2009) Intralaboratory and interlaboratory evaluation of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. Mutat Res 673(2):100-8. doi: 10.1016/j.mrgentox.2008.12.003. https://www.sciencedirect.com/science/article/pii/S1383571809000060?via%3Dihub.

- Kidd D, Phillips S, Chirom T, Mason N, Smith R, Saul J, Whitwell J, Clements J (2021) The 3D reconstructed skin micronucleus assay: considerations for optimal protocol design. Mutagenesis 36(1):37-49. doi: 10.1093/mutage/gez037. https://academic.oup.com/mutage/article/36/1/37/5650914?login=true.

- Kirkland D, Aardema M, Henderson L, 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. doi: 10.1016/j.mrgentox.2005.02.004. https://www.sciencedirect.com/science/article/abs/pii/S1383571805000835?via%3Dihub.

- Kirkland D, Kasper P, Martus HJ, Mueller L, van Benthem J, Madia F, Corvi R (2016) Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. Mutat Res 795:7-30. doi: 10.1016/j.mrgentox.2015.10.006.

- Madia F, Corvi R (2020) EURL ECVAM genotoxicity and carcinogenicity consolidated database of Ames negative chemicals. European Commission, Joint Research Centre (JRC) [Dataset]

PID: http://data.europa.eu/89h/38701804-bc00-43c1-8af1-fe2d5265e8d7.

- Mun GC, Aardema MJ, Hu T, Barnett B, Kaluzhny Y, Klausner M, Karetsky V, Dahl EL, Curren RD (2009) Further development of the EpiDerm 3D reconstructed human skin micronucleus (RSMN) assay. Mutat Res 673(2):92-9. doi: 10.1016/j.mrgentox.2008.12.004. https://www.sciencedirect.com/science/article/pii/S1383571809000059?via%3Dihub.
- OECD (2016a) TG 474: Mammalian Erythrocyte Micronucleus Test, adopted 29 July 2016. https://www.oecd-ilibrary.org/environment/test-no-474-mammalian-erythrocyte-micronucleus-test_9789264264762-en.
- OECD (2016b) TG 489: *In Vivo* Mammalian Alkaline Comet Assay, adopted 29 July 2016. https://www.oecd-ilibrary.org/environment/test-no-489-in-vivo-mammalian-alkaline-comet-assay_9789264264885-en.
- OECD (2016c) TG 475: Mammalian Bone Marrow Chromosomal Aberration Test, adopted 29 July 2016. https://www.oecd-ilibrary.org/environment/test-no-475-mammalian-bone-marrow-chromosomal-aberration-test_9789264264786-en.
- OECD (2022) TG 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays, adopted 30 June 2022. https://www.oecd-ilibrary.org/environment/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_9789264203907-en.
- OECD (2023) TG 487: *In Vitro* Mammalian Cell Micronucleus Test, adopted 4 July 2023. https://www.oecd-ilibrary.org/environment/test-no-487-in-vitro-mammalian-cell-micronucleus-test_9789264264861-en.
- Pfuhler S, Downs TR, Hewitt NJ, Hoffmann S, Mun GC, Ouedraogo G, Roy S, Curren RD, Aardema MJ (2021) 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. Mutagenesis 36(1):1-17. doi: 10.1093/mutage/geaa035. https://academic.oup.com/mutage/article/36/1/1/6129202?login=true.
- Pfuhler S, van Benthem J, Curren R, Doak SH, Dusinska M, Hayashi M, Heflich RH, Kidd D, Kirkland D, Luan Y, Ouedraogo G, Reisinger K, Sofuni T, van Acker F, Yang Y, Corvi R (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 Genet Toxicol Environ Mutagen 850-851:503135. doi: 10.1016/j.mrgentox.2020.503135. https://www.sciencedirect.com/science/article/pii/S138357182030005X?via%3Dihub.
- Sasaki JC, Allemang A, Bryce SM, Custer L, Dearfield KL, Dietz Y, Elhajouji A, Escobar PA, Fornace AJ Jr, Froetschl R, Galloway S, Hemmann U, Hendriks G, Li HH, Luijten M, Ouedraogo G, Peel L, Pfuhler S, Roberts DJ, Thybaud V, van Benthem J, Yauk CL, Schuler M (2020) Application of the adverse outcome pathway framework to genotoxic modes of action. Environ Mol Mutagen 61(1):114-134. doi: 10.1002/em.22339. https://onlinelibrary.wiley.com/doi/10.1002/em.22339.
- SCCS (2023) Scientific Committee on Consumer Safety. Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation - 12th revision, from pages 74-75. Adopted May 15, 2023. SCCS/1647/22. https://health.ec.europa.eu/publications/sccs-notes-guidance-testing-cosmetic-ingredients-and-their-safety-evaluation-12th-revision_en.

- Sommer S, Buraczewska I, Kruszewski M (2020) Micronucleus assay: The state of art, and future directions. Int J Mol Sci 21(4):1534. doi: 10.3390/ijms21041534.
- Thakkar Y, Moustakas H, Aardema M, Roy S, Pfuhler S, Api AM (2022) Use of the EpiDerm™ 3D reconstructed skin micronucleus assay for fragrance materials. Mutagenesis 37(2):89-111. doi: 10.1093/mutage/geab040.
- 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. doi: 10.1159/000358272. https://karger.com/spp/article/27/5/263/295819/Dermal-Xenobiotic-Metabolism-A-Comparison-between.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
Joint-research-centre.ec.europa.eu

Publications Office of the European Union