Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2026)13

Unclassified

English - Or. English

29 June 2026

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

# Peer Review Report of the Epi2sensA for Skin Sensitisation

## OECD Series on Testing and Assessment

E-mail: ehscont@oecd.org.

JT03590121

This work was approved and declassified by the Chemicals and Biotechnology Committee on [02/07/2026].

**Please cite this publication as:**

OECD (2026), *Peer Review Report of the Epi2sensA for Skin Sensitisation*, OECD Series on Testing and Assessment, No. 435, OECD Environment, Health and Safety, Paris, link to ONE M&P document https://one.oecd.org/document/ENV/CBC/MONO(2026)13/en/pdf

**Contact us**

**OECD Environment Directorate,**
**Environment, Health and Safety Division**
**2 rue André-Pascal**
**75775 Paris Cedex 16**
**France**

**E-mail: ehscont@oecd.org**

# About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 countries in North and South America, Europe and the Asia and Pacific region, as well as the European Union, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several Partner countries and from interested international organisations attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (https://www.oecd.org/en/topics/chemical-safety-and-biosafety.html).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank, Basel, Rotterdam and Stockholm Conventions and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

# Foreword

This document contains the Peer review report of the EpiDerm-Epidermal Sensitisation Assay (Epi2SensA), a me-too method of the Epidermal Sensitisation Assay (EpiSensA), an in vitro method for identifying the skin sensitisation potential of chemicals.

The scientific validity of the Epi2SensA was assessed by an international Peer Review Panel between August and November 2025. The peer review report supports the development of the corresponding test method for inclusion in Appendix IC of Test Guideline 442D for in vitro skin sensitisation assays, which assesses the Adverse Outcome Pathway Key Event on Keratinocyte activation. The Peer review report and the Validation report of the Epi2SensA were circulated for comments to the Working Party of the National Coordinators of the Test Guidelines Programme (WNT) in December 2025. The WNT approved the peer review report at its 38th meeting in April 2026. The WNT also approved the Validation report of the Epi2SensA.

This report is published under the responsibility of the Chemicals and Biotechnology Committee.

# Report of the Peer Review Panel (PRP)

on

The catch-up validation study of Epi2SensA

**Table of Contents**



## 1. Executive Summary

The Epi2SensA method is a proposed "me-too" assay of the original Epidermal Sensitisation Assay (EpiSensA). It was evaluated against the Performance Standards (PS) outlined in OECD Series on Testing and Assessment No. 396 (Appendix V), intended to facilitate the validation of new or modified EpiSensA test method.

While Epi2SensA retains the core scientific principles and biomarker genes of the VRM, several model-dependent modifications to components and decision criteria were required to meet performance standards. Such adjustments are not uncommon and have been accepted in other OECD-validated methods; therefore, the Peer-Review Panel considered the changes justified and proceeded with the evaluation.

The modified Epi2SensA method, which uses the EpiDerm™ model (MatTek Corp., Ashland, MA) and incorporates adapted acceptance and prediction criteria, successfully met the PS for reliability and predictive capacity. The average WLR (83.3%) and the BLR (85%) both exceeded the minimum requirement of 80%. Although one of the laboratories participating to validation study (BRT) did not individually meet the 80% WLR threshold (75.0%), the overall performance of the method was deemed acceptable by the PRP.

Based on the 20 chemicals tested, the assessment of Predictive Capacity showed an average Sensitivity of 88.1%, Specificity of 88.9%, and Accuracy of 88.3%, all of which met or exceeded the minimum PS acceptance criteria ($\geq$ 85% for sensitivity and accuracy, and $\geq$ 65% for specificity). The PRP concluded that the method's overall performance meets the performance standards and is sufficiently robust.

## 2. Peer Review Panel Composition

The scientific validity of the Epi2SensA was assessed by an international Peer Review Panel between August 2025 and November 2025. The peer-review was organized by the National Coordinators for the OECD TG Program from France and Japan. The international PRP

comprised nominations from international validation bodies via OECD. Each PRP member provided a declaration of interest as to ensure no conflicts of interest existed (note: the declaration of interests can be made available upon request).

Emanuela Corsini, PhD (PRP Chair) – Università degli Studi di Milano, Italy
Hans Raabe, MS – Institute for In Vitro Sciences, Inc., USA
Verma Rajeshwar, PhD – FDA, USA*
Marla Dubau, PhD – German Federal Institute for Risk Assessment, Germany
Li Huan, PhD – Temasek Polytechnic, Singapore
* Due to the current U.S. government shutdown, Dr Rajeshwar was unable to access or respond to emails. This has impacted his ability to contribute to the finalisation of this review.

The validation study report was made available to the PRP at the end of August 2025. Overall, and a total of two teleconferences took place during the peer-review of Epi2SensA as described below.
19 Sept. 2025: PRP discussed the deviations from the original protocol, considered them acceptable, thus agreed to proceed with evaluation.
15 Oct. 2025: PRP discussed the first draft.

The second draft of the PRP evaluation of the Epi2SensA test method was prepared by 20 Oct. 2025, agreed by e-mail exchanges and finalised by November.

## 3. Study objective and purpose of the Epi2SensA method

The Epidermal Sensitization Assay (EpiSensA) as the Epi2SensA is a gene expression–based test method employing reconstructed human epidermis (RhE) to predict the skin-sensitizing potential of chemicals. Owing to the presence of a stratified stratum corneum in the RhE model, the assay is suitable for evaluating a broad range of test substances, including lipophilic compounds as well as pre-haptens and pro-haptens. This structural feature enables exposure conditions that closely mimic those of in vivo assays and actual human contact.

## 4. Comparison of Epi2SensA with the essential test method components and performance, as described in Appendix V for the assessment of the similar EpiSensA test method (Emanuela)

The Epi2SensA method is a proposed "me-too" assay of the original Epidermal Sensitisation Assay (EpiSensA), referred to as the Validated Reference Method (VRM). Epi2SensA was assessed against the Performance Standards (PS) outlined in the OECD Series on Testing and Assessment No. 396 (Appendix V).
The comparison reveals that while Epi2SensA maintains the core scientific principle and marker genes of the VRM, it required several model-dependent modifications to its essential

components and decision criteria to successfully meet the required performance standards. This is not unique to this assay as examples of adjustments can be found in other OECD accepted methods, and therefore, the Peer-Review Panel considered changes acceptable and proceeded with the evaluation. The Peer-Review Panel also noted that in the first publication of test developer (Saito et al., 2013) EPI-200 (24 well format) and EPI-296 (96 well format) of EpiDerm™ were used as experimental model, and the concurrent upregulation of five genes, namely ATF3, DNAJB4, GCLM, HSPA6 and HSPH1) was considered in the classification of a chemical as skin sensitisers.

At the end of the experimental study, the statistician conducted the data analysis using the original acceptance criteria, resulting in an average WLR of 63.9%, a BLR of 70%, and averages of 66.6% specificity, 90.5% sensitivity, and 83.3% accuracy. In the Table below (Table 11 of the Validation Report), the comparison of the WLR, BLR and predictive capacity according to the original analysis and according to the modified acceptance criteria, with green highlighted cells indicating met acceptance criteria are reported.

| | acceptance criterion | Mattek | | Eurofins | | BRT | | Average original | Average modified |
|---|---|---|---|---|---|---|---|---|---|
| | | original | modified | original | modified | original | modified | | |
| WLR | ≥ 80% | 66,7% (8/12) | 91,7% (11/12) | 58,3% (8/12) | 83,3% (10/12) | 66,7% (8/12) | 75,0% (9/12) | 63,9% | 83,3% |
| BLR | ≥ 80% | original: 70% | | | modified: 85% | | | 70% | 85% |
| Specificity | ≥ 65% | 66,7% (4/6) | 100% (6/6) | 83,3% (5/6) | 83,3% (5/6) | 50,0% (3/6) | 83,3% (5/6) | 66,6% | 88,9% |
| Sensitivity | ≥ 85% | 92,9% (13/14) | 92,9% (13/14) | 85,7% (12/14) | 78,6% (11/14) | 92,9% (13/14) | 92,9% (13/14) | 90,5% | 88.1% |
| Accuracy | ≥ 85% | 85,0% (17/20) | 95,0% (18/20) | 85,0% (17/20) | 80,0% (16/20) | 80,0% (16/20) | 90,0% (18/20) | 83,3% | 88,3% |
| Balanced accuracy | | 82.89% | 96,5% | 84.5% | 81% | 71.5% | 88,1% | 78.6% | 88.5% |

**Comparison of Essential Test Method Components**

The Essential Test Method Components define the core structural, functional, and procedural elements required for a similar assay.

| **Essential Test Method Component** | **VRM (EpiSensA) Requirement [Appendix V]** | **Epi2SensA (EpiDerm Implementation [Report]** | **Key Difference & Rationale** |
|---|---|---|---|
| **Reconstructed Human Epidermis (RhE) Model** | LabCyte EPI-MODEL24. | **EpiDerm (EPI-200)**. | **Different model.** This was the fundamental difference requiring numerous subsequent protocol adjustments. |
| **Marker Genes** | Quantifies expression of ATF3, GCLM, DNAJB4, and IL-8. | Quantifies expression of ATF3, GCLM, DNAJB4, and IL-8. | **Identical.** Both methods target the same four mechanistically relevant genes associated with keratinocyte activation. |

| | | | |
|---|---|---|---|
| **Gene Cut-off Values** | ATF3 > 15-fold; GCLM > 2-fold; DNAJB4 > 2-fold; IL-8 > 4-fold. | ATF3 > 15-fold; GCLM > 2-fold; DNAJB4 > 2-fold; IL-8 > 4-fold. | **Identical.** The gene-specific induction thresholds were retained. |
| **Cytotoxicity Viability Threshold** | Must maintain cell viability **> 80%** for acceptable test concentration results. | Must maintain cell viability **> 60%** for acceptable test concentration results. | **Modified Criterion.** The threshold was reduced from 80% to 60% based on preliminary data showing the LDH assay overestimated cytotoxicity for EpiDerm compared to the MTT assay, and to enhance test sensitivity. |
| **Prediction Model (Positive Result)** | Prediction is positive if **at least one** marker gene exceeds its cut-off (Imax) at an accepted concentration. | Prediction is positive if **at least two** marker genes exceed their respective cut-off values (Imax) at an accepted concentration. | **Modified Criterion.** The requirement was increased to two positive genes to enhance robustness, a modification common when using different tissue models in me-too validation. |
| **Exposure Time** | 6 hours. | 1 hour topical exposure followed by a 5-hour post-incubation period. | **Modified Procedure.** The exposure duration was shortened to 1 hour to reduce unexpected cytotoxicity observed with the EpiDerm model, while maintaining the 6-hour time point for gene expression measurement. |
| **Application Volume** | 5 µl applied to the epidermis surface. | 10 µl applied to the epidermis surface. | **Modified Procedure.** The volume was doubled because the surface area of the EpiDerm model (0.63 $cm^2$ is roughly double that of the LabCyte model (0.32 $cm^2$) thus maintaining a similar volume/surface ratio. |
| **Positive Control (Clotrimazole)** | 0.78%(w/v). | **1.56% (w/v)**. | **Modified Procedure.** The concentration was increased to ensure the run acceptance criteria were consistently met for ATF3 and IL-8 fold induction, as 0.78% |

| | | | |
|---|---|---|---|
| | | | led to a 40% failure rate in preliminary Epi2SensA runs. |
| **Killed Control Method** | 10 μl of 10% Triton X-100 applied topically. | 50 μl of 10% Triton X-100 applied in the culture medium. | **Modified Procedure.** Changed the volume and application location to ensure complete tissue death and maximum LDH release for the EpiDerm model. |

**Comparison of Defined Reliability and Accuracy Performance**

The Performance Standards require minimum concordance and accuracy values for a similar method to be deemed acceptable.

Initial Performance Analysis (Using VRM Criteria)

When the Epi2SensA data were analysed using the original VRM acceptance criteria (80% viability, one positive gene), the method failed to meet the criteria for reliability:

| Performance Metric | PS Minimum Requirement | VRM Actual Performance (Based on Validation Dataset) | Epi2SensA Original Analysis (Average) | Status |
|---|---|---|---|---|
| **WLR** | 80.0% | 88.8% | 63.9% | **FAILED** |
| **BLR** | 80.0% | 94.1% | 70% | **FAILED** |
| **Accuracy** | 85.0% | 85.0% | 83.3% | FAILED (by average) |
| **Sensitivity** | 85.0% | 92.9% | 90.5% | Met |
| **Specificity** | 65.0% | 66.7% | 66.6% | Met |

**Modified Performance Analysis (Using Adapted Criteria)**

Following the failure of the initial analysis, the VMT implemented the two major modifications (60% viability threshold and two positive genes) and re-analysed the data. The adapted criteria resulted in a substantial improvement in performance, enabling the method to meet the required PS criteria.

| Performance Metric | PS Minimum Requirement | VRM Actual Performance | Epi2SensA Modified Analysis (Average) | Status |
|---|---|---|---|---|
| **WLR** | 80.0% | 88.8% | **83.3%** | **Met** |
| **BLR** | 80.0% | 94.1% | **85%** | **Met** |
| **Accuracy** | 85.0% | 85.0% | **88.3%** | **Met** |
| **Sensitivity** | 85.0% | 92.9% | 88.1% | **Met** |

| | | | | |
|---|---|---|---|---|
| **Specificity** | 65.0% | 66.7% | **88.9%** | **Met** |

**Conclusion on Performance**

With the necessary model-dependent adaptations to the acceptance criteria and prediction model, the Epi2SensA method demonstrated overall acceptable reliability and relevance. The average Within-Laboratory Reproducibility (WLR, 83.3%), Between-Laboratory Reproducibility (BLR, 85%), Sensitivity (88.1%), and Accuracy (88.3%) all met or exceeded the minimum acceptance criteria defined in the Performance Standards for similar methods.

While not changing the overall performance of the Epi2SensA, the peer-review panel considered that the justification provided for the inconsistency of the results reported on Table 8 (page 14 of the Validation report) not sufficient. The VMT's argument is that the misclassification of Methyl heptine carbonate as a non-sensitizer (NS) in BRT Run 6 (sample C342-2) was an isolated error likely caused by handling or concentration issues, rather than a systemic failure of the method. The justification relies heavily on the *likelihood* of an error rather than confirmed documentation of one. This raised some questions:

- What specific documentation (e.g., laboratory notebook entries, temperature logs, material balance checks, or external visual evidence of the volatile sample preparation) was reviewed by the VMT to conclude that a handling or concentration error occurred, beyond the resultant non-concordant data point itself?
- Was the incorrect grouping assignment itself considered a protocol deviation or a source of variability in the validation, given that relying on low concentrations (3.13% down to 0.39% for C342-2) for a substance that should be tested much higher (100% down to 6.25% for C176-0, B130-0, A115-0) increased the risk of misclassification, regardless of the handling error?
- Did the VMT investigate whether the BRT technician's inexperience with RT-PCR or the handling of volatile chemicals was explicitly addressed or mitigated between the interim Phase 2 analysis (March 28, 2025) and the completion of Run 6 (which produced the outlier result)?

This could clarify if the technical adjustment period was insufficient to prevent this specific type of error.

## 5. Biological and mechanistic relevance

In accordance with the OECD adverse outcome pathway (AOP) for skin sensitization (OECD, 2014), four key events describe the progression from the molecular interaction to the adverse outcome. Key Event 2 (KE2) corresponds to the keratinocyte activation, involving the induction of inflammatory responses and the expression of antioxidant response element (ARE)-regulated genes. The Epi2SensA assay addresses this step by measuring gene expression changes in a RhE that reproduces the stratified architecture and partial metabolic competence of the skin.

The Epi2SensA assay quantifies the expression of four mechanistically anchored genes, *ATF3*, *IL-8*, *GCLM*, and *DNAJB4*, that have been previously established as biomarkers in the validated reference method (EpiSensA). Each of these genes has been demonstrated to be associated with a distinct biological pathway that is pertinent to KE2. *ATF3* is as a central regulator of the cellular adaptive response, modulating cytokine and chemokine expression through ATP- and NF-κB-dependent signaling (Ohara et al., 2010; Gilchrist et al., 2006). IL-8 is a pro-inflammatory cytokine that functions as a chemotactic mediator for neutrophils. Its expression is regulated by ATP–P2X7 receptor activation and downstream signaling through the p38 MAPK pathway. (Saito et al., 2017; Hoffmann et al., 2002). *GCLM* encodes the modifier subunit of glutamate-cysteine ligase, the rate-limiting enzyme for glutathione biosynthesis. Its expression is regulated by the nuclear factor E2-related factor 2 (Nrf2)/ARE and activator protein-1 (AP-1) pathways, thus linking oxidative stress to cytoprotective responses (Saito et al., 2017; Lu et al., 2013). *DNAJB4*, a heat-shock protein co-chaperone, involved in stabilizing protein folding under oxidative stress (Qiu et al., 2006), is regulated by the Nrf2/ARE- and the heat shock transcription factor-1/heat shock factor response element (HSF-1/HSE) pathway (Satoh et al., 2011; Mizumachi et al., 2023). Collectively, these markers signify cytoprotective, antioxidant, and inflammatory processes that are characteristic of keratinocyte activation upon sensitizer exposure

The biological plausibility of this marker panel of Epi2SensA is supported by the established literature showing consistent induction of these genes by sensitizing chemicals but not by non-sensitizers. These pathways are recognized as integral to the cellular response preceding dendritic-cell activation and immune priming, thus linking the assay endpoints to the skin sensitization AOP. The RhE model further enhances physiological relevance by enabling topical exposure and partial metabolic conversion of test substances (Luu-The et al., 2009, Götz et al., 2012a, Götz et al., 2012b), extending applicability to pre- and pro-haptens.

A key strength of the Epi2SensA assay is its use of mechanistically relevant biomarkers in a RhE model, offering a physiologically meaningful representation of keratinocyte activation during KE2. Moreover, the reliance on gene expression changes in well-characterized stress and inflammatory pathways provides a transparent mechanistic link to sensitizer activity. Nevertheless, as with other KE2-based in-vitro assays, Epi2SensA does not encompass downstream immunological events such as dendritic-cell activation or T-cell proliferation. The sensitivity of the method to weak sensitizers or substances requiring extensive metabolic activation remains to be confirmed, and the restricted biomarker panel may not cover all chemical mechanisms.

**Conclusion on biological and mechanistic relevance**

The PRP considered that the Epi2SensA assay demonstrates an adequate and scientifically well-justified biological and mechanistic relevance. Its use of a RhE model and four mechanistically anchored gene markers provides a coherent representation of keratinocyte activation consistent with KE2 of the skin sensitization AOP.

## 6. Test Method Protocol

A detailed test method protocol in the form of a Standard Operating Procedure (SOP) document is available, written in such a way that a naïve facility should be able to easily pick up the test method for their use. The Epi2SensA SOP was based upon the SOP of the validated reference method (VRM), with modifications initially made by the lead laboratory, MatTek Corporation, during SOP development to address the morphometric differences between the VRM's LabCyte EPI-MODEL-24 and the MatTek Corporation EpiDerm™ EPI-200 reconstructed human epidermis (RhE) model, such as total tissue surface area and relative barrier functions. Specific differences in the SOP between Epi2SensA and the VRM were focused upon the exposure kinetics, including dosing volume, exposure duration, and post-treatment incubation times, while other modifications to improve method performance were also made. The Epi2SensA SOP included a larger dose volume of 10 µL per tissue, as compared to the 5 µL dose volume in the VRM, which was modified given the approximately 2-fold larger surface area of the EpiDerm™ model. As a result of misclassifications and excessive cytotoxic effects early in method development the 6-hour continuous test chemical exposure was modified to a 1-hour test chemical exposure, followed by a rinsing step, and a 5-hour post-exposure expression incubation. This modification was justifiable and consistent with the exposure kinetics of the EpiDerm™ skin irritation test procedure in OECD Test Guideline 439, and the 6-hour total timeframe has been shown to allow for upregulation of gene induction for measuring skin sensitization not only in the VRM, but also in another RhE-based test method (SENS-IS, Immunosearch, Grasse, France).

Additional SOP modifications were made during method development to facilitate the LDH release cytotoxicity endpoint procedures. It was recognized that the procedures used in the VRM to lyse the tissues designated for use as the killed control (used to establish the 100% LDH release value) were not effective in fully releasing preformed LDH from the killed control, thereby resulting in inaccurate assessments of cytotoxicity induced by test substances, which would impact the selection of exposures appropriate for gene expression analyses. Specifically, it was demonstrated that the application of 10 µL of 10% Triton-X topically onto the apical surface of the killed control tissue for 6 hours was not fully effective at lysing the tissue. Therefore, the Epi2SensA procedure was modified to apply the 10% Triton-X into the medium under the killed control tissue for a baso-lateral exposure, thereby bypassing the tissue barrier function, and to increase the volume of 10% Triton-X from 10 µL to 50 µL. Initial experimental data showed this modification to be effective in releasing essentially all of the tissue's LDH. Lastly, the concentration of the positive control, Clotrimazole was increased from the concentration of 0.78% cited in the VRM, to 1.56% to ensure meeting acceptance criteria. The Validation Study Report presents data from 10 runs comparing responses for Clotrimazole at both 0.78% and 1.56% wherein increasing the dose concentration resulted in more consistent acceptable responses.

The Validation Study Report highlighted that the method was transferable from the lead laboratory to three naïve laboratories (Food and Drug Safety Center Hatano Research Institute (FDSC), Eurofins BioPharma Product Testing, and Burleson Research Technologies)

with some minor adjustments to the protocol, which were captured in subsequent versions. Participating laboratories underwent initial training to learn the method and subsequently evaluate their proficiency by testing the 2 positive controls and 4 non-coded chemicals. Whereas the Validation Study Report presents that 5 chemicals were tested in the training phase, Appendix I of the report only identifies 4 chemicals. Those 4 chemicals did not belong to the chemical set used to formally evaluate the me-too test method.

The SOP was further revised to its final form during the conduct of the testing of the coded chemicals. Specifically, as a result of an interim analysis of test method performance by the VMT of data from the first 2 runs, the acceptance criteria for tissue viability as well as the prediction model were modified to better retrofit to the data to improve test method sensitivity and specificity. The acceptance criteria for tissue viability was modified by lowering the cell viability threshold from 80% to 60%, thereby including more gene expression data for analyses. This change may be justified in concert with the revised LDH lysis killed control procedures, as well as differences in chemical permeation kinetics in the RhE tissues used in the Epi2SensA SOP relative to the VRM. To enhance specificity, the prediction model was modified to require at least two significant gene expression changes for a positive prediction, rather than just one, as in the VRM. These specific SOP modifications, while highly impacting, did not change physical parameters or procedures during the conduct of the validation study, and thus are not considered to undermine the goals to evaluate test method performance. Furthermore, these changes do not appear to have any influence upon a Study Director's decision process during the conduct of a study and hence would not have impacted the study execution decision processes by the Study Directors during the validation. Accordingly, the Peer Review Panel does not consider these SOP changes to have adversely impacted the goals of the test method validation. Indeed, in the classical test method validation paradigm, such method optimizations would be expected to be done during test method prevalidation trials. It is also noteworthy that the thresholds for positive predictions for each of the 4 genes has not changed from the thresholds specified for the VRM, showing a high level of mechanistic relevance between the Epi2SensA test method and the VRM.

## 7. Appropriateness of the validation study management and conduct

The primary goals of this validation study were to assess the test method definition, relevance, transferability and within- and between-laboratory reproducibility (reliability) as a similar (or colloquially, a "me-too") test method to the fully adopted validated reference method (VRM) EpiSensA.

The Validation Study Report presents on the overall validation study plan, the composition of the Validation Management Team (VMT), the selection and training of the participating laboratories, adherence to GLP principles, test chemical selection, coding and distribution, and guidance and communications during the validation study.

- Study Plan: the overall study plan included an initial training and demonstration of proficiency phase (Phase 1) to determine whether the SOP was sufficiently robust for

transferability to 3 naïve labs, followed by Phase 2 for evaluating test method performance with the 20 reference chemicals presented in the Performance Standards for the VRM.

- At least 3 labs, each testing the 20 blind-coded chemicals
- For resource, cost and time savings a subset of 12 of the 20 chemicals were tested in three valid independent runs to assess within lab reproducibility (WLR). The remaining were run in single runs to support between lab reproducibility (BLR). The VMT consulted the OECD Expert Group on this accepted study design.

NOTE: that due to workload constraints, one of the 3 naïve labs was not able to commit to completing the full set of coded chemicals, so the lead laboratory joined as a participating lab to support the determinations of WLR and BLR.

- Phase 2 was also divided into two portions as a result of an interim data analysis of the data from the first two runs. This did not seem to have any adverse or influencing impact on the outcome of the validation study.

- Validation Management Team: The Validation Study Report presents the composition of the VMT, which at the core includes 5 experts in test method validation activities. However, the VMT also includes up to three representatives from the 4 testing laboratories), including the Study Directors, whose roles are identified as Observer.

NOTE: it is not clear from the Validation Study Report what the role of the Observer is within the VMT, since traditionally Study Directors as well as all staff under their supervision are not included within the VMT, to avoid any uncontrolled communications from influencing the Study Directors' decisions and actions on the conduct of the test method. Furthermore, once the training and transferability activities have been completed representatives from the testing laboratories are precluded from having any communications amongst them related to the validation study. Thus, in the absence of any clear statement on communications expectations and restrictions, it is not clear to the Peer Review Panel whether any interlaboratory communications occurred that could have influenced interpretation of the SOP, impacted various decisions on execution work conduct, biased expectations of results, or caused other undue impacts.

- Adherence to GLP Principles: The study was not conducted in full GLP compliance by any of the laboratories, and accordingly in the absence of an independent auditing oversight, any claims to adherence to GLP principles simply cannot be verified. The Validation Study Report presents that all of the laboratories performed all procedures in accordance with GLP principles, including, but not limited to, the use of standard operating procedures, proper data recording and record keeping. Rather than making the aforementioned statements, it is more useful and informational to state which elements of full GLP compliance were not implemented or executed, to provide greater clarity into potential vulnerabilities – and thereby instil greater confidence. Nonetheless, all of the participating

laboratories have a history of routinely conducting toxicology and research studies both non-GLP and in full GLP compliance.

- Whereas assuring that the wet lab activities, equipment and facility performance, reagent preparation, and staff training can be relatively easily verified by the written records and documentation, the integrity and authenticity of the electronic data and records from initial readings on the PCR thermocycler through to the excel spreadsheets used to analyse the data are notably less visible and notably more vulnerable to inadvertent errors and omissions. The Epi2SensA and VRM methods rely heavily on high density data such that in the absence of a thorough data audit, the Peer Review Panel cites this potential vulnerability.

- Chemical Acquisition, Coding, Distribution and Handling: The 20 recommended reference chemicals cited in the VRM Performance Standards were acquired by the Chemical Management representative on the VMT (VitroScreen). Although the key representative originally cited on the VMT resigned their role, the function of acquiring, coding, labelling and distributing the reference chemicals was carried out.
  - A unique blind code system was employed which precluded the ability of individuals within a laboratory to decipher the chemical identities, and which precluded individuals between laboratories to compare test outcomes. This is a basic requirement for blind coding chemicals.
  - Furthermore, for those test chemicals used for WLR analyses, the code system precluded revealing the identity of chemicals to be used in repeat runs, which minimizes the ability of laboratory staff to utilize knowledge of outcomes from previous runs in the execution of subsequent runs.
  - The blind coded reference chemicals were delivered to the laboratories along with sealed envelopes each containing MSDS for each reference chemical. The envelopes only included the blind code, to identify the specific reference chemical.

NOTE: it is not clear from the Validation Study Report how the 10 envelopes were labelled for the chemicals that had unique codes for each trial.

- The Validation Study Report presents that guidance was provided on securing and maintaining the MSDS envelopes, as well as downstream documentation to show the codes were not revealed/compromised.
- To minimize costs associated with the dose range finding experiments, only 10 of the 20 reference chemicals were to be tested in the initial cytotoxicity test. For the remaining 10 reference chemicals the recommended doses for use in the definitive gene expression tests were presented to the testing facilities. There were three different dose ranges presented, representative of low, mid and high cytotoxicities. A priori knowledge of these groupings does not impart the ability to definitively identify the individual chemicals, nor does it impart expectations for gene expression

changes, thus precluding undue influence or bias on the validation outcome.

- Each laboratory did receive information on the physical state, sample weight or volume, and storage requirements from VitroScreen. It was not stated in the Validation Study Report whether laboratory staff were instructed on or utilized universal precautions and personal protective equipment.

- Transferability and Operator Training
  - Early misclassifications at Eurofins were attributed to inexperienced RT-PCR technicians and protocol-harmonization meetings midway through the trial—indicating insufficient pre-validation training.
  - Eurofins used a distinct RNA extraction kit and the technician was also in experienced with the PCR method.

Note: GLP requires that all personnel be appropriately educated, trained, and experienced for their assigned functions. However, despite the report's statement that participating laboratories operated in accordance with GLP principles, Eurofins appears not to have met this requirement.

## 8. Recommendations and Further Considerations

Borderline range specifications are now being required of all test methods in OECD TG 442D, particularly with respect to gaining confidence in the use of resulting data in defined approaches presented in OECD GL 497. Accordingly, it is recommended that the test method developers begin to analyse borderline ranges for the individual gene expression thresholds, utilizing robust data sources. These data should include the data generated from the coded chemicals from the test method validation, and may be supplemented with data from peer reviewed publications and from audited/vetted data derived from implementation of the test method according to the final version of the Epi2SensA SOP.

The Validation Study Report presents the composition of the VMT, which at the core includes 5 experts in test method validation activities. However, the VMT also includes up to three representatives from the 4 testing laboratories), including the Study Directors, whose roles are identified as Observer. It is recommended that additional information be provided which details the Observer role, and what expectations and restrictions for communications were implemented among the VMT and participating laboratory staff personnel.

Whereas OECD GD 34 on test method validation does not require validation studies to be conducted in full GLP compliance, as non-animal test methods have progressed from relatively simple endpoints and data analyses to complex high data content/black box analyses, there is a growing need for greater transparency in the documentation of data capture, maintenance, handling and analyses to impart universal confidence in test methods

depending on such data rich endpoints. Independent auditing of raw and analysed data used to support validation study findings will need to be conducted to the same degree as for GLP-compliant regulatory studies, and there are growing expectations that such independent data audits be part of any validation data package.

Although the training and transfer of the Epi2SensA method were conducted systematically through online meetings, video tutorials, and three training runs coordinated by the lead laboratory, further consideration should be given to documenting training effectiveness and associated workload. Two technicians per laboratory were trained, and all participating laboratories were ultimately deemed proficient. However, variability observed during early runs suggests that differences in prior RT-PCR experience and hands-on familiarity with RhE models influenced the performance. Future validation activities would benefit from clearer documentation of training duration, evaluation of trainee proficiency, and associated workload to support confidence in method transferability and reproducibility.

## 9. REFERENCES

Gilchrist M, Thorsson V, Li B, Rust AG, Korb M, Roach JC, Kennedy K, Hai T, Bolouri H, Aderem A. Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. Nature. 2006;441(7090):173-178. doi:10.1038/nature04768.

Götz C, Pfeiffer R, Tigges J, Blatz V, Jäckh C, Freytag EM, Fabian E, Landsiedel R, Merk HF, Krutmann J, Edwards RJ, Pease C, Goebel C, Hewitt N, Fritsche E. Xenobiotic metabolism capacities of human skin in comparison with a 3D epidermis model and keratinocyte-based cell culture as in vitro alternatives for chemical testing: activating enzymes (Phase I). Exp Dermatol. 2012 May;21(5):358–363. doi:10.1111/j.1600-0625.2012.01486.x. PMID: 22509833.

Götz C, Pfeiffer R, Tigges J, Ruwiedel K, Hübenthal U, Merk HF, Krutmann J, Edwards RJ, Abel J, Pease C, Goebel C, Hewitt N, Fritsche E. Xenobiotic metabolism capacities of human skin in comparison with a 3D-epidermis model and keratinocyte-based cell culture as in vitro alternatives for chemical testing: phase II enzymes. Exp Dermatol. 2012 May;21(5):364–369. doi:10.1111/j.1600-0625.2012.01478.x. PMID: 22509834.

Hoffmann E, Dittrich-Breiholz O, Holtmann H, Kracht M. Multiple control of interleukin-8 gene expression. J Leukoc Biol. 2002 Nov;72(5):847-55. doi:10.1189/jlb.72.5.847. PMID: 12429706.

Lu SC. Glutathione synthesis. Biochim Biophys Acta. 2013 May;1830(5):3143-3153. doi:10.1016/j.bbagen.2012.09.008.

Luu-The V, Duche D, Ferraris C, Meunier JR, Leclaire J, Labrie F. Expression profiles of phases 1 and 2 metabolizing enzymes in human skin and the reconstructed skin models Episkin and full thickness model from Episkin. J Steroid Biochem Mol Biol. 2009 Sep;116(3-5):178-86. doi: 10.1016/j.jsbmb.2009.05.011. Epub 2009 May 29. PMID: 19482084.

Mizumachi H, Suzuki S, Sakuma M, Natsui M, Imai N, Miyazawa M. Reconstructed human epidermis-based testing strategy of skin sensitization potential and potency classification using epidermal sensitization assay and in silico data. J Appl Toxicol. 2024 Mar;44(3):415–427. doi:10.1002/jat.4551. Epub 2023 Oct 17. PMID: 37846211

OECD (2014), The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, OECD Series on Testing and Assessment, No. 168, OECD Publishing, Paris, https://doi.org/10.1787/9789264221444-en.

Ohara H, Saito R, Hirakawa S, Shimada M, Mano N, Okuyama R, Aiba S. Gene expression profiling defines the role of ATP-exposed keratinocytes in skin inflammation. J Dermatol Sci. 2010;58(2):143-151. doi:10.1016/j.jdermsci.2010.02.007.

Qiu XB, Shao YM, Miao S, Wang L. The diversity of the DnaJ/Hsp40 family: the crucial partners for Hsp70 chaperones. Cell Mol Life Sci. 2006 Nov;63(22):2560–2570. doi:10.1007/s00018-006-6192-6. PMID: 16952052.

Saito K, Nukada Y, Takenouchi O, Miyazawa M, Sakaguchi H, Nishiyama N. Development of a new in vitro skin sensitization assay (Epidermal Sensitization Assay; EpiSensA) using reconstructed human epidermis. Toxicol In Vitro. 2013 Dec;27(8):2213-24. doi: 10.1016/j.tiv.2013.08.007. Epub 2013 Aug 30. PMID: 23999411.

Saito K, Takenouchi O, Nukada Y, Miyazawa M, Sakaguchi H. An in vitro skin sensitization assay termed EpiSensA for broad sets of chemicals including lipophilic chemicals and pre/pro-haptens. Toxicol In Vitro. 2017;40:11-25. doi:10.1016/j.tiv.2016.12.005.

Satoh T, Rezaie T, Seki M, Sunico CR, Tabuchi T, Kitagawa T, Yanagitai M, Senzaki M, Kosegawa C, Taira H, McKercher SR, Hoffman JK, Roth GP, Lipton SA. Dual neuroprotective pathways of a pro-electrophilic compound via HSF-1-activated heat-shock proteins and Nrf2-activated phase 2 antioxidant response enzymes. J Neurochem. 2011 Nov;119(3):569–578. doi:10.1111/j.1471-4159.2011.07449.x. Epub 2011 Sep 21. PMID: 21883218.