Epidermal Sensitization Assay (EpiSensA)
JaCVAM Validation Study Report

Version 5

September 2022

Contents

Summary 5
Background 7
Management of the Study 10
1. Study objectives 10
2. Study Plan 10
2-1. Structure of the study 10
2-2. Validation Management Team 11
2-3. Participating Laboratories 12
3. Test Design 13
4. Selection of Test Chemicals 14
5. Acquisition, coding, and distribution 15
5-1. Chemical Acquisition, Coding, and Distribution 15
5-2. Handling 15
6. Data management 15
7. Statistical Analysis of Test Data 16
7-1. Data Analysis 16
7-2. Evaluation Criteria 16
Test Definition 18
1. Intended purpose of the test method 18
2. The evidence demonstrating the need of the test method 18
3. Development of the test method 18
4. Scientific basis: biological and mechanistic relevance 23
5. SOP (Standard Operating Procedure) 25
6. Test Definition: Conclusions of the Validation Management Team 26
Transferability 28
1. General aspects 28
2. Training 28
3. Transfer of the test method to participating laboratories 29
4. Results of technical transfer 30
5. Modifications made to the original SOP 34
6. Conclusions of the VMT 36
Within-Laboratory Reproducibility 37
1. Phase I-A 37
1-1. LION 37

1-2. KOSÉ 39
1-3. FDSC 41
1-4. Revision of the SOP and other consideration 43
2. Phase I-B 46
2-1. LION 46
2-2. KOSÉ 48
2-3. FDSC 49
2-4. Revision of the SOP and other consideration 50
3. Phase I-C 51
3-1. LION 51
3-2. KOSÉ 52
3-3. FDSC 54
3-4. Revision of the SOP and other consideration 55
4. Overall result of WLR and other considerations 56
4-1. WLR 56
4-2. Discussion of discordant results 56
4-3. Consideration around the positive controls 57
5. Conclusion 60
Between-Laboratory Reproducibility 61
1. Between-Laboratory Reproducibility (concordance in prediction) 61
1-1. Between-laboratory reproducibility of the 12 test chemicals used in Phase II 61
1-2. Between-laboratory reproducibility of the 27 test chemicals used in Phases I and II 64
2. Other considerations 65
2-1. Discussion of non-concordant-BLR chemicals 65
2-2. Consideration around the positive controls 67
3. Conclusion 74
Quality Check 75
1. Chemical distribution 75
1-1. Chemical Acquisition, Coding and Distribution 75
1-2. Handling 75
2. Quality assurance 75
Predictive Capacity 77
1. Predictive capacity for LLNA data 77
2. Predictive capacity for human data 79
3. Predictive capacity, sorted into three categories of sensitization potency 81
4. Comparison to the results of lead laboratory 81

5. Test chemicals with results that were false negatives 82
6. Test chemicals with results that were false positives 85
7. Conclusion 90
Applicability domain 91
1. Technical limitations 91
1-1. Solubility of test chemicals 91
2. Summary of the SOP revision 91
3. Limitations on applicability 92
2-1. Endogenous control gene 92
2-2. Restrictions on cell viability assays 93
2-3. Restriction on RNA extraction 93
2-4. Inherent limitation of a threshold-based prediction model 93
2-5. Pre/pro-haptens 93
VMT overall conclusions and recommendations 94
1. Overall conclusions 94
2. Recommendations 95
Acknowledgement 97
References 98

- Appendix 1 Study plan
- Appendix 2 QC report
- Appendix 3 Statistical report
- Appendix 4 Chemical selection
- Appendix 5 EpiSensA dataset including 136 chemicals
- Appendix 6 Comparability of equipment
- Appendix 7 EpiSensA SOP
- Appendix 8 Articles of EpiSensA
- Appendix 9 Minutes of VMT meeting
- Appendix 10 Reactivity Domain and Physicochemical Property of EpiSensA Dataset
- Appendix 11 The detail of the revision of the SOP in the Phase I-C
- Appendix 12 Results of EpiSensA and existing test methods at 27 validation chemicals

Summary

This report presents the results of a study of the epidermal sensitization assay (EpiSensA) that was carried out at three participating laboratories in order to validate the assay's within- and between-laboratory reproducibility, and to evaluate its predictive capacity as compared to that of the Local Lymph Node Assay (LLNA).

The assessment of skin sensitization potential is an important part of the overall safety assessment of both existing and of newly developed test chemicals. Historically, animal tests such as the Guinea Pig Maximization Test (GPMT) or the LLNA have been the most employed methods of testing for skin sensitization potential. But the limitations of these animal-based methods as well as the increased awareness of animal welfare issues has in recent years prompted the development, validation, use, and regulatory acceptance of non-animal testing methods based upon a knowledge of adverse outcome pathways (AOP). Validated non-animal methods for the assessment of skin sensitization potential include the Direct Peptide Reactivity Assay (DPRA) (key event; KE1), KeratinoSens™ and LuSens (KE2), as well as the human Cell Line Activation Test (h-CLAT), U-SENS™, and the IL-8 Luc assay (KE3).

Following a modular approach of applying the European Center for the Validation of Alternative Methods (ECVAM) principles on test validity (OECD Series on Testing and Assessment, Number 34, 2005; Hartung *et al.*, 2004), the Validation Management Team (VMT) has assessed modules 1 through 4 (test definition, within-laboratory reproducibility, transferability, and between-laboratory reproducibility), and has also used the validation study results in order to assess modules 5 and 6 (predictive capacity and applicability domain).

Regarding the transferability phase of the study, three sensitizers and one non-sensitizer were correctly judged in all main studies at all participating laboratories with good reproducibility. The transfer of EpiSensA from the lead laboratory to the three participating laboratories was completed successfully and to the satisfaction of both the lead laboratory and the VMT.

As part of the Phase I study, each of the three participating laboratories performed three experiments of identical sets of 15 coded test chemicals in order to evaluate the within-laboratory reproducibility. The results, based on concordance, were 93.3% (14/15), 93.3% (14/15), and 86.7% (13/15). When the results from the 15 test chemicals of Phase I were combined with those of Phase II – in which each participating laboratory performed one experiment of identical sets of an additional 12 test chemicals – the

between-laboratory reproducibility for 27 test chemicals was 88.9%. The VMT considered the EpiSensA to be very reproducible within and between laboratories, satisfying the performance criteria for this class of test method.

Furthermore, taking into account chemicals tested outside this validation study, the predictive performance of the assay for identifying skin sensitizers also has proved promising (accuracy and balanced accuracy vs. LLNA was 82.4% and 76.9% at 136 chemicals, respectively).

Background

Allergic contact dermatitis (ACD) resulting from skin sensitization is one of the most common skin diseases contributing to a decreased quality of life for patients. Skin sensitization is a significant toxicological endpoint for cosmetic ingredients, and the assessment of skin sensitization has traditionally been dependent upon animal tests such as the local lymph node assay (LLNA; Organization for Economic Cooperation and Development [OECD], 2010) and the guinea pig maximization test (GPMT; OECD, 1992); for both of which the OECD framework has already provided test guidelines. In the regulatory context, only data deriving from these animal experiments has been conventionally acceptable for the assessment of the skin sensitizing potential of chemicals. On the other hand, the development of alternative test methods for identifying the skin sensitization potential of chemicals has recently been activated as a result of ethical concerns as well as the testing and concomitant marketing ban imposed by the European Union (Regulation (EC) No 1223/2009 of the European Parliament and of the Council, 2009).

While skin sensitization consists of highly complex multifactorial events, the chemical and biological key events related to the induction of skin sensitization have been described in detail. The OECD has already reported on the AOP for skin sensitization (OECD, 2012) by defining the key events in the sensitization process. In brief, the key events include: (i) the binding of small reactive chemicals (called "haptens") to skin proteins (key event 1; KE1); (ii) the inflammatory response and the induction of cytoprotective gene pathways in keratinocytes (KE2); (iii) the induction of surface molecules and cytokines, resulting in dendritic cell mobilization and migration to the lymph nodes (KE3); and (iv) the presentation of the resulting histocompatibility complexes to naive T cells by the aforementioned dendritic cells, with a subsequent differentiation and proliferation of the activated T cells (KE4). OECD has released several test guidelines focusing on these key events in order to assess the skin sensitization potential of various chemicals. Most recently, the Direct Peptide Reactivity Assay (DPRA), the Amino acid Derivative Reactivity Assay (ADRA), and the kinetic Direct Peptide Reactivity Assay (kDPRA) (KE1), KeratinoSens™ and LuSens (KE2; mainly addressing the cytoprotective gene pathways in keratinocytes but not their inflammatory response), as well as the human Cell Line Activation Test (h-CLAT), U-SENS™, and IL-8 Luc assay (KE3) have all been adopted as OECD Test Guidelines (OECD, 2018a; 2018b; 2019).

However, these tests have some shared limitations. For example, the cell-based assays in general have less predictivity for chemicals with high logKow values as a result of their aqueous-phase system. When a test chemical is not soluble or fails to form a stable and uniform dispersion, the testing cannot be conducted. For the KeratinoSens™ and the LuSens methods, chemicals with logKow > 7 may not be adequately dissolved in the culture media and, as a result, negative results need to be considered with caution if cytotoxicity is observed (OECD, 2018a). For the h-CLAT, negative results for test chemicals with logKow > 3.5 should not be considered as valid due to solubility issues (OECD, 2018b). It is also necessary to consider that the assays have a weakness in detecting pre- and pro-haptens that need to undergo abiological oxidation and metabolic conversion in order to become sensitizing haptens (Aptula *et al.*, 2007). Both the DPRA and ADRA have no metabolic capacity, and thus they are not expected to be able to correctly classify pro-haptens (OECD, 2019). The KeratinoSens™ and other cell line-based assays present with fewer problems in that respect, due to their limited oxidative or metabolic activity (OECD, 2018a; 2018b). Testing strategies called Defined Approaches (DAs) that combine different information from multiple AOP-based assays, which are used in a specific combination, and resulting data are interpreted using a fixed data interpretation procedure, have been recently developed and evaluated (OECD, 2016a; 2016b); however, all these DAs still have some limitations due to applicability domain being used for each assay. Therefore, the chemical space for which each assay is applicable needs to be thoroughly expanded.

To this end, we have focused on a reconstructed human epidermis (RhE) model consisting of normal human-derived epidermal keratinocytes (NHEKs). The epidermis is fully stratified and covered by a stratum corneum. Like in the case of an animal test, various kinds of test chemicals (including lipophilic ones) can be directly applied on the stratum corneum by dissolving the first in both hydrophilic and lipophilic vehicles. Moreover, RhE models have been reported to show a metabolic capacity similar to that of an animal or a human skin (Oesch *et al.*, 2018, Tokudome *et al.*, 2015), suggesting that pre- and pro-haptens might be correctly assessed by using an RhE model. Recently, some RhE model-based skin sensitization assays have been reported, such as those of the RhE/IL-18 assay (Gibbs et al., 2013; Andres *et al.*, 2017), the SenCeeTox® assay (McKim *et al.*, 2012), and the SENS-IS (Cottrez *et al.*, 2016). The SENS-IS, for example, that measures the expression of 61 genes whose products are primarily related to redox, inflammation, and tissue repair processes, has shown 93% and 100% accuracy when applied for 28 lipophilic chemicals with logKow > 3.5 and 14 pre- and pro-haptens, respectively (Cottrez *et al.*, 2016). A recent study using different datasets for 126

chemicals has reported that the SENS-IS is capable of providing an 80.2% accuracy, a 70.2% balanced accuracy, a 90.4% sensitivity, and a 50% specificity in predicting the LLNA results; rates that are of higher predictivity to those generated by the KeratinoSens™ (67.2% accuracy, 67.1% balanced accuracy, 67.4% sensitivity, 66.7% specificity) (Hoffmann *et al.*, 2018).

The epidermal sensitization assay (EpiSensA) is an *in vitro* test method addressing KE2. EpiSensA is based on the commercially-available human skin model LabCyte EPI-MODEL 24 in which gene expressions are measured for four mechanistically-relevant markers, namely: (i) the encoding activating transcription factor 3 (*ATF3*), (ii) the glutamate-cysteine ligase, modifier subunit (*GCLM*), (iii) the DnaJ (Hsp40) homolog, subfamily B, member 4 (*DNAJB4*), and (iv) interleukin-8 (*IL-8*). The expression of these genes reflects the keratinocyte response to the early phase of skin sensitization, including: (i) the induction of cyto-protective gene pathways (e.g., the antioxidant response element (ARE)-dependent pathways) and (ii) the inflammatory response (e.g., the induction of inflammatory cytokines) (OECD, 2012).

This study reports on the validation of EpiSensA in predicting the skin sensitization potential of chemicals.

Management of the Study

1. Study objectives

The EpiSensA validation study was launched in 2018. The primary study objective was to evaluate the technical transferability and reliability (based on within- and between-laboratory reproducibility) of EpiSensA, with the anticipation of its future use in combinatorial approaches with other alternative test methods as replacements of the animal tests.

A secondary objective of this study was to obtain the experimental data generated during the reliability evaluation. The data were then used in order to assess the performance of EpiSensA in distinguishing skin sensitizers (Category 1) from non-sensitizers (No Category), as defined by the UN Globally Harmonized System (GHS) classification, and in the labelling of substances for skin sensitization.

This report presents the results of the testing performed at three independent laboratories, primarily in order to validate the transferability and reliability of EpiSensA.

2. Study Plan

The Study Plan was drafted, approved and issued by the Validation Management Team (VMT) prior to the start of the testing. The final version is annexed to this report (see Appendix 1).

2-1. Structure of the study

The EpiSensA validation study was organized in accordance with the modular approach described in the Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (OECD Series on Testing and Assessment, Number 34, 2005). Moreover, the study was designed in order to generate the information related to modules 1-4 (1: test definition; 2: within-laboratory reproducibility; 3: transferability; 4: between-laboratory reproducibility) of the European Center for the Validation of Alternative Methods (ECVAM) modular approach to validation (Hartung *et al.*, 2004). Additionally, the data obtained during the validation study were also used for the purpose of evaluating module 5 (predictive capacity) and module 6 (applicability).

The study was entirely coordinated by the Japanese Center for the Validation of Alternative Methods (JaCVAM) in consultation with the VMT with regard to the study design, the test chemical selection, and the standard operating procedure (SOP). Figure 1 illustrates how the validation study was organized in terms of the management followed, the test methods included, the participating laboratories, the selection, the coding and the supply of the test chemicals, as well as the data collection and their statistical analysis. Full details on the management, sponsorship, coordination, timings, responsibility, and overall set-up of the study are provided in Appendix 1.

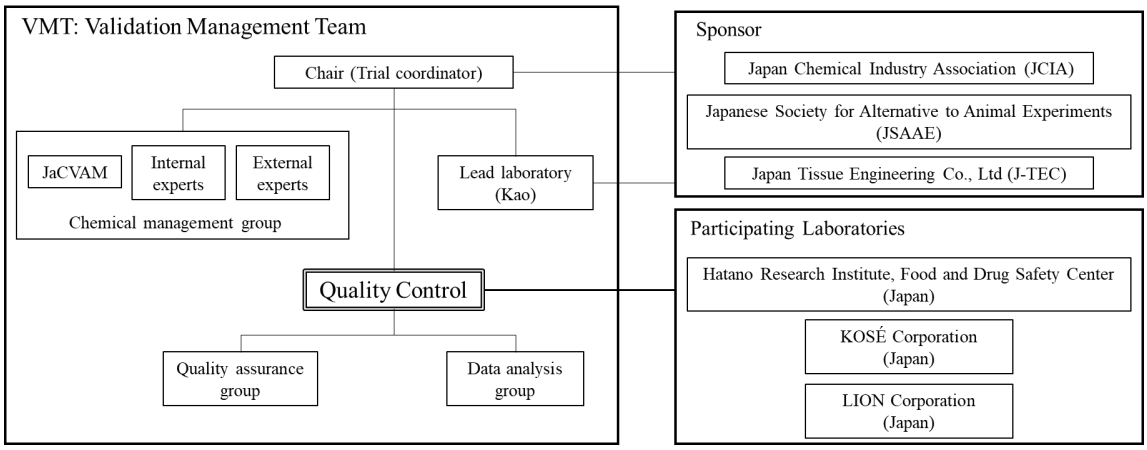


Figure 1. EpiSensA validation study structure and organization.

2-2. Validation Management Team

The VMT: (i) encompasses collective expertise on the test, (ii) specifies the study objectives, the design and the required performance criteria, and (iii) administers the testing, the evaluation and the presentation of the test results.

Dr. Hajime Kojima, who is the JaCVAM representative and has experience as a trial coordinator for the preparation of OECD test guidelines, served as chair of the VMT. The other members of the VMT are shown below.

Validation Management Team

Chair (Trial Coordinator)	Hajime Kojima (JaCVAM representative, NIHS)
Statistician	Takashi Sozu (Tokyo University of Science)
Internal Expert	Masahiro Takeyoshi (CERI)
	Takao Ashikaga (JaCVAM, NIHS)
External Expert	David Basketter (DABMEB Consultancy Ltd)
	Chantra Eskes (SeCAM)
	Sebastian Hoffmann (seh consulting + services)
	David M. Lehmann (U.S. EPA)
	Tae Sung Kim (KoCVAM)
Lead Laboratory	Masaaki Miyazawa (Kao Corporation)
	Hideyuki Mizumachi (Kao Corporation)

2-3. Participating Laboratories

The following three laboratories participated in the EpiSensA validation study:

Laboratory 1
Hatano Research Institute,
Food and Drug Safety Center,
Kanagawa, Japan
Mika Watanabe, Hajime Sui, Tomoko Shindo (study directors)

Laboratory 2
KOSÉ Corporation,
Safety and Analytical Research Laboratories,
Tokyo, Japan
Noriyasu Imai, Makoto Mizuno (study directors)

Laboratory 3
Lion Corporation
Safety Research Science Laboratory,
Kanagawa, Japan
Shin-ichi Watanabe (study director)

3. Test Design

Reference document: Appendix 1

The validation study was implemented in two independent stages:

Transferability stage: Training of the participating laboratories with the transfer of the technology and the expertise required for them to perform the test method, and to confirm the transferability of the test method.

Reliability stage: Evaluation of the protocol performance based on the results of testing performed under blind conditions at all the participating laboratories; there were two phases to this stage of the study (see below).

The test chemicals used during the Transferability stage were selected by the lead laboratory. Initial training was conducted with four uncoded test chemicals containing three sensitizers and one non-sensitizer, for which unambiguous results were expected. In addition, two of the four tested chemicals were potential candidates for the new positive control. The results were assessed by the lead laboratory and the VMT before the participating laboratories started to prepare for the Reliability stage.

The number and type of test chemicals used during the Reliability stage were selected by the VMT, by taking into account the testing period of the EpiSensA validation study (from October 2018 to July 2021) and the required performance criteria. As a result, the Phase I study evaluating the within- and between-laboratory reproducibility was conducted using three replicate sets of 15 coded test chemicals, while the Phase II study evaluating the between-laboratory reproducibility was conducted using one set of 12 coded test chemicals. In addition, the Phase I study was divided into three parts, namely Phase I-A, I-B, and I-C.

A breakdown of the test chemicals used in Phases I and II reads as follows:

Phase I-A: 4 sensitizers, 1 non-sensitizer
Phase I-B: 2 sensitizers, 3 non-sensitizers
Phase I-C: 4 sensitizers, 1 non-sensitizer
Phase II: 8 sensitizers, 4 non-sensitizers

4. Selection of Test Chemicals

Reference document: Appendix 4

The test chemicals for the validation study were selected by the VMT after taking into account published *in vivo* skin sensitization studies and previous validation studies of *in vitro* alternative methods for predicting the skin sensitization potential. Specific consideration was given to the following aspects:

- the information on mode/site of action,
- the quality and quantity of the available reference data (*in vivo* and *in vitro* testing),
- the availability of high-quality data deriving from not only animal but also (if available) from human studies,
- the coverage of a range of relevant chemical and product classes,
- information on interspecies variations (e.g., interspecies variabilities regarding the assimilation of chemicals, metabolism, etc.),
- the coverage of a range of toxic effects and/or sensitizing potencies,
- information about indicating pre- and pro-haptens,
- the physical and chemical properties of the tested compounds (and their suitability for experimental use as implied by their CAS No.),
- single chemical entities or formulations known to be of high purity,
- commercial availability, as well as
- cost.

At the first chemical selection meeting, the VMT has prepared a primary dataset by identifying and collecting several existing lists of potential skin sensitizers. This list included chemicals that had already been used in other validation studies, were referred in other test methods and/or in OECD TG 429, and were recommended for use in testing by EURL ECVAM. Moreover, the VMT performed an extensive literature search in order to ensure that all chemicals fulfilled the above criteria.

The diversity of potencies (Category 1A, Category 1B and non-sensitizer) was also emphasized at the chemical selection meeting. Additionally, the VMT determined that at least a third of the test chemicals selected should be non-sensitizers.

5. Acquisition, coding, and distribution

5-1. Chemical Acquisition, Coding, and Distribution

The evaluation of interlaboratory transferability was performed using the results of tests conducted at the participating laboratories with non-coded chemicals. On the other hand, the evaluation of the within- and the between-laboratory reproducibility, as well as of the predictive capacity were made with using coded chemicals. The coding was supervised by JaCVAM. In addition, JaCVAM was responsible for acquisition, coding, and distributing all test and control chemicals used in the validation study.

5-2. Handling

Each participating laboratory was provided with the essential information about the test chemicals (e.g., their physical state, their sample weight or volume, and their storage instructions) by JaCVAM. Each laboratory was responsible for storing the respective chemicals in accordance with the storage instructions, and separately received the sealed safety information, including Safety Data Sheets (SDSs) describing their hazard identification, their exposure control, and the personal protection required for each chemical. The test chemicals were delivered directly to the study directors of each laboratory. The SDSs were accessed only in the event of accident, and the information were disclosed only to those who had to know.

6. Data management

Prior to the start of the study, a standard data reporting template was prepared and distributed to the participating laboratories. The template was made by the lead laboratory.

The template contained a formula that was checked by the Data Analysis Group in a documented exercise prior to the template distribution. The participating laboratories were then asked to use the template during the Transferability stage in order to compile and communicate their results to the lead laboratories. This way, they were able to familiarize themselves with the template, and this ensured that the instructions included in the templates were clear and understood before the start of the blind testing. The formula-containing template was protected from editing in order to prevent any accidental change.

All the results obtained from the Reliability stage were then submitted by the participating laboratories directly to the Data Analysis Group and the Quality Assurance

Group by e-mail. The quality check focused on the acceptance criteria for the run and for each of the chemical data in order to ensure that the results were valid and to confirm the correct judgment of prediction.

For the undertaking of the statistical analysis, a summary template with tables and graphs was designed by the statistician, and the results were transferred onto this template by JaCVAM. The preparation of this summary template contained internal checks that ensured that no transcription errors were made in the transfer of the results. As an additional check, the final conclusions/outcomes for each chemical were then compared to the conclusions/outcomes appearing in the reports sent by the laboratories.

7. Statistical Analysis of Test Data

7-1. Data Analysis

Only test results that fulfilled the acceptance criteria described in the SOP were subjected to statistical analysis, although records were provided of all test results that contained the data that did not met the acceptance criteria. Failed runs and experiments were also used in order to assess their occurrence and frequency.

A detailed statistical analysis method was produced and agreed at the first VMT meeting prior to the start of the testing phases. The reliability of the test method was assessed by calculating the within- and the between-laboratory reproducibility. The statistical analysis on the reproducibility focused on the concordance of classification, sensitizer (S) versus non-sensitizers (NS). Additionally, the data of the positive control were also analyzed in order to assess the frequency of their acceptance.

7-2. Evaluation Criteria

In order to assess whether the objectives of this validation study had been met, it was necessary to define in advance the minimum performance criteria required for the assessment of the test method performance. When defining the criteria, the following factors were considered as important:

1. the background and objectives of the validation study,
2. the realistically expected performance of the test method based on an *in vitro* test and the standards of performance that have been considered as acceptable in previous validation studies,

3. the proposed use of the *in vitro* test method (i.e., to be used in combination with other alternative test methods), as well as
4. the statistical power of the design of the validation study.

As a result, the minimum performance criteria for this validation study were set at 85% for the within-laboratory reproducibility, and at 80% for the between-laboratory reproducibility.

Test Definition

See also the EpiSensA SOP (Appendix 7) and the articles of EpiSensA (Appendix 8)

1. Intended purpose of the test method

The EpiSensA is a potential partial replacement *in vitro* test method designed to be part of a non-animal test battery or of an integrated testing strategy for the assessment of the skin sensitization potential of chemicals. The purpose of this testing method is to contribute to the reduction of the number of animals used for the undertaking of skin sensitization testing as well as for the replacement of the currently employed regulatory *in vivo* tests for skin sensitization hazard classification and labelling (e.g., OECD TG 406, TG 429).

2. The evidence demonstrating the need of the test method

The EpiSensA test method is particularly important in view of the European Union ban on *in vivo* testing of cosmetic ingredients and products implemented by the Seventh Amendment to the Cosmetics Directive as well as by the European Regulation on Registration, Evaluation, and Authorization of Chemicals (REACH). The success of the EpiSensA validation study would support its implementation in the assessment of the skin sensitization potential of chemicals, and it would reduce or eliminate the number of animal testing required for this same reason.

In Japan, Pharmaceuticals and Medical Devices Agency (PMDA) has recently accepted a "bottom-up 3 out of 3" which is an integrated strategy with three alternative in vitro methods (i.e. DPRA, KeratinoSens™, and h-CLAT) (Amber *et al.*, 2018). Since EpiSensA is applicable to a broad range of chemicals, including pre/pro-hapten and lipophilic chemicals, EpiSensA could also be a useful assessment tool from the regulatory perspective in Japan.

3. Development of the test method

EpiSensA was developed by the Kao Corporation.

Extensive work in order to develop and optimize the EpiSensA protocol had been undertaken since 2011. The AOP KE2 of skin sensitization involved keratinocytes being stimulated in order to activate their antioxidant response genes and to release their danger signals (e.g., pro-inflammatory cytokines or ATP). At the preliminary study, microarray analysis and subsequent real-time PCR analysis were performed in order to select the skin sensitization marker genes based on their ability to: (i) be characterized by significantly marked induction by sensitizers, but not non-sensitizers, (ii) be relevant to the keratinocyte response to skin sensitization, and (iii) be reproducible through both the microarray and the real-time PCR results. These analyses revealed that the expression of six genes (*GCLM*, *DNAJB4*, *IL-8*, *ATF3*, *HSPA6*, and *HSPH1*) could provide reliable biomarkers for the discrimination of sensitizers from non-sensitizers (Saito *et al.*, 2013; 2017). Subsequently, 16 chemicals (8 sensitizers, 4 non-sensitizers, and 4 pre-/pro-haptens) were used in order to assess the predictive performance of the six marker gene candidates (Saito *et al.*, 2013; 2017). These 16 chemicals have been recommended by the ECVAM as reference chemicals for the development of *in vitro* skin sensitization tests (Casati *et al.*, 2009). Finally, four genes (*ATF3, IL-8*, *GCLM*, and *DNAJB4*) were selected as EpiSensA marker genes.

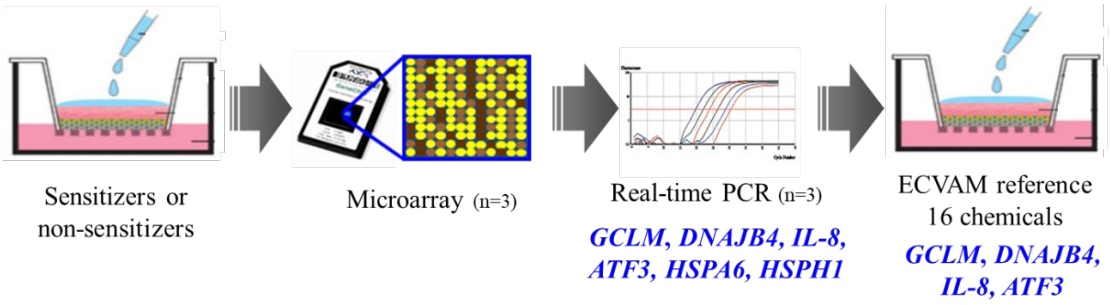


Figure 2. Selection flow of the marker genes.

Based on 72 test chemicals containing 54 sensitizers and 18 non-sensitizers, the cut-off values of the respective marker genes were decided as the predictive performance of EpiSensA was maximized (Saito *et al.*, 2017). In addition, the cut-off values were reconsidered after using 136 chemicals that had LLNA EC3 values (101 sensitizers and 35 non-sensitizers in LLNA; Appendix 5). Figure 3 illustrates the relationship between maximal fold induction (Imax) of four marker genes at 136 chemicals and LLNA EC3 values. Figure 3 also shows that it is difficult to discriminate sensitizers from non-sensitizers by employing only one marker gene, and that the combination of marker genes

may be needed in order to improve the assay's sensitivity. By focusing on non-sensitizers with exceeding cut-off values of at least one of the marker genes, 9 out of the 12 non-sensitizers have shown a positive gene expression at more than 1 of the marker genes. It is notable that the majority of these non-sensitizers may possess very weak sensitization potential (Table 1). Furthermore, three other chemicals with no clear information regarding their sensitization potential (diethyl phthalate, diethyl toluamide, and sulfanilamide) have induced a markedly increase of multiple marker genes, and it may not be possible to distinguish these three chemicals from sensitizers. Therefore, the cut-off value should be set without considering them. Moreover, Table 2 summarizes the sensitizers rated as positive in only one marker gene. Table 2 clearly shows that each marker gene has extreme or strong sensitizers that can be detected by only one marker gene (e.g., *ATF3*: lauryl gallate and propyl gallate; *GCLM*: chlorothalonil; *DNAJB4*: hexyl salicylate; *IL-8*: benzoyl peroxide). For that reason, the respective cut-off values should be less than the Imax value of these extreme/strong sensitizers. Finally, from these results, we were able to conclude that the cut-off values that have been set by Saito *et al.* (2017) (namely, *ATF3*: 15-fold, *GCLM*: 2-fold, *DNAJB4*: 2-fold, and *IL-8*: 4-fold) were also appropriate based on our current assessment of 136 chemicals.

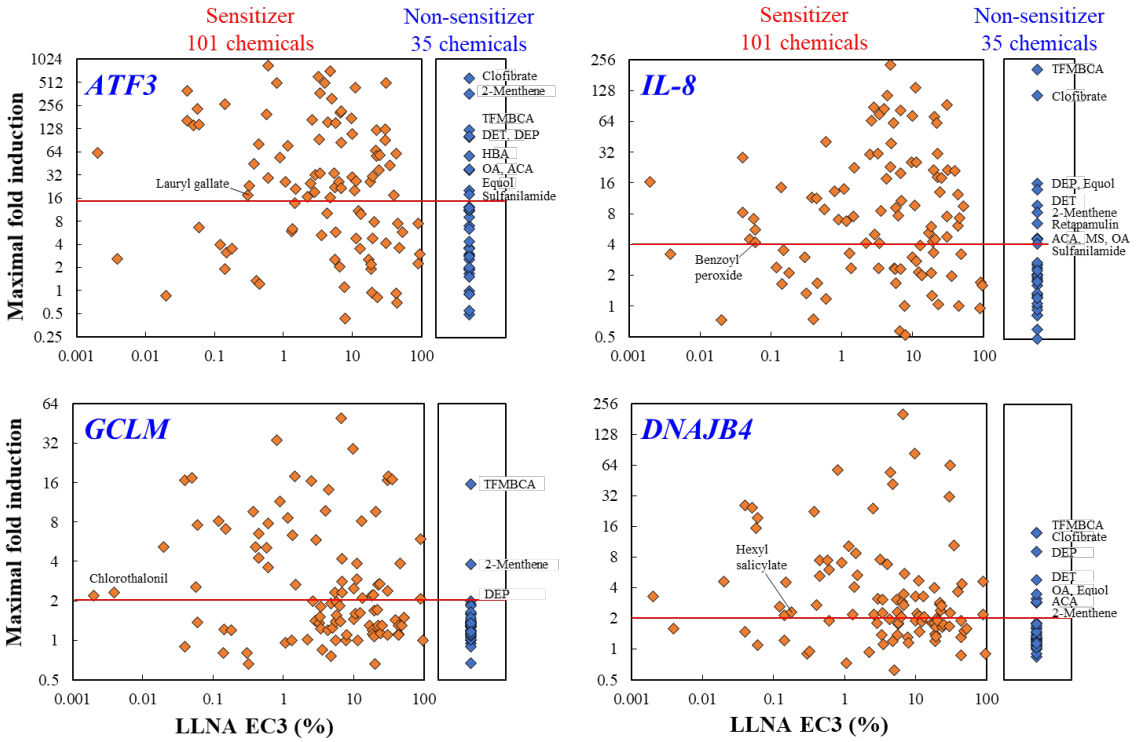


Figure 3. The relationship between the Imax of four marker genes at the presence of each of the tested 136 chemicals and the LLNA EC3 values (2-menthene: (+)-trans-p-menth-2-ene; TFMBCA: 4'-trifluoromethylbiphenyl-4-carbaldehyde; DET: diethyl toluamide; DEP: diethyl phthalate; HBA: 4-hydroxybenzoic acid; OA: octanoic acid, ACA; α-amylcinnamyl alcohol, MS; methyl salicylate).

Table 1. LLNA negative chemicals rated as EpiSensA positive for more than one of the marker genes.

Chemical	CAS No.	Imax				Comments
		ATF3	*GCLM*	*DNAJB4*	*IL-8*	
α-Amylcinnamyl alcohol	101-85-9	**37.5**	1.4	**2.9**	**6.1**	Borderline induction in LLNA (SI=2.9 at 10% exposure conc.)
Clofibrate	637-07-0	**574.4**	1.2	**14.0**	**206.4**	Borderline induction in LLNA (SI=2.9 at 50% exposure conc.)
Diethyl phthalate	84-66-2	**98.7**	**2.0**	**9.1**	**23.8**	The is no clear information of skin sensitization hazard.
Diethyl toluamide	134-62-3	**121.9**	1.9	**4.8**	**13.7**	The is no clear information of skin sensitization hazard.
Equol	531-95-3	**20.4**	1.8	**3.1**	**20.4**	All of DPRA, KeratinoSens™ and h-CLAT also judged as positive.
(+)-trans-p-Menth-2-ene	5113-93-9	**359.4**	**3.8**	**2.9**	**11.4**	All of DPRA, KeratinoSens™ and h-CLAT also judged as positive.
Octanoic acid	124-07-2	**38.1**	1.1	**3.5**	**5.9**	Some fatty acids (oleic acid, linoleic acid etc.) reported as positive in LLNA.
Sulfanilamide	63-74-1	**18.2**	1.3	0.9	**5.3**	The is no clear information of skin sensitization hazard.
4'-Trifluoromethylbiphenyl-4-carbaldehyde	90035-34-0	**101.9**	**15.7**	**14.1**	**393.0**	Borderline induction in LLNA (SI=2.6 at 10% exposure conc.)

Table 2. Sensitizers rated as positive in only one marker gene.

	CAS No.	LogKow	EC3(%)	Imax			
				ATF3	*GCLM*	*DNAJB4*	*IL-8*
Lauryl gallate	1166-52-5	6.21	**0.30**	**17.6**	0.8	0.9	3.0
Propyl gallate	121-79-9	1.79	**0.32**	**22.9**	0.7	0.9	1.3
Diethylenetriamine	111-40-0	-2.13	3.3	**92.7**	1.5	1.4	2.4
Embramine hydrochloride	13977-28-1	4.45	5.5	**22.1**	1.0	1.4	2.3
Chlorothalonil	1897-45-6	3.66	**0.004**	2.6	**2.3**	1.6	3.2
Trifluralin	1582-09-8	5.31	11	4.7	**3.9**	1.5	2.8
Hexyl salicylate	6259-76-3	5.06	**0.18**	3.5	1.2	**2.3**	2.1
1-Naphthol	90-15-3	2.69	1.3	5.9	1.0	**2.2**	3.3
2-Methoxy-4-methylphenol	93-51-6	1.88	5.6	2.5	1.4	**2.3**	2.3
4-Chloroaniline	106-47-8	1.72	6.5	2.0	1.4	**2.9**	0.6
Benzoyl peroxide	94-36-0	3.43	**0.06**	6.6	1.4	1.1	**4.2**
Bourgenol	18127-01-0	3.94	4.3	10.1	1.2	2.0	**17.6**
Benzyl benzoate	120-51-4	3.54	17	2.5	1.3	1.8	**5.2**
1,1,3-Trimethyl-3-phenylindane	3910-35-8	5.91	43	0.9	1.1	0.9	**6.1**
Limonene	5989-27-5	4.83	52.5	5.8	1.5	1.6	**9.5**

Regarding the predictive performance based on 136 chemicals obtained by lead laboratory, when compared to LLNA results, the sensitivity, specificity, accuracy, and balanced accuracy were 88.1% (89/101), 65.7% (23/35), 82.4% (112/136), and 76.9%, respectively (Appendix 5). The predictive performance was comparable to that of existing methods [e.g., the sensitivity, specificity, accuracy, and balanced accuracy of h-CLAT were 93% (94/101), 66% (27/41), 85% (121/142), and 79.5%, respectively (OECD, 2018b)]. In addition, the EpiSensA had a sensitivity of 82.7%, a specificity of 64.7%, an accuracy of 78.3%, and a balanced accuracy of 73.7% for 69 lipophilic chemicals (Appendix 5). Regarding pre/pro-haptens, the EpiSensA demonstrated 97.3% sensitivity, and almost pre/pro-haptens could be detected as positive (Appendix 5). Furthermore, the predictive performance of EpiSensA for GHS sub-categorization (Cat.1A, Cat.1B, NC), which was based on LLNA results, was also evaluated. The prediction model for potency classification is described in Appendix 5. Results demonstrated that the EpiSensA had a potency accuracy of 71.3%. In addition, the predictive performance for distinguishing GHS Cat.1A from not GHS Cat.1A (Cat.1B and NC) was also compared to that of kDPRA. It was confirmed that the sensitivity (the proportion of GHS Cat.1A chemicals that are correctly classified) of EpiSensA (73.7%) was comparable to that of kDPRA (84.2%) on the same set of 72 chemicals (Appendix 5). Moreover, the potency predictive performance of EpiSensA for human results [based on GHS sub-categorization described in Annex 2 of OECD Guideline No.497 (OECD, 2021) or Basketter's classification (Basketter et al., 2014)] was also evaluated and compared to that of LLNA and kDPRA.

It was confirmed that the potency accuracy of EpiSensA for trinary sub-categorization vs. human results was 64.4% and comparable to that of LLNA (64.4%), and the sensitivity of EpiSensA (66.7%) for distinguishing GHS Cat.1A from not GHS Cat.1A was also comparable to that of kDPRA (58.3%) (Appendix 5).

Regarding the applicability domain, the chemical space of the 136 dataset have been checked by lead laboratory. The chemical reactivity domains and physicochemical properties of 136 chemicals in EpiSensA dataset were compared to those of 196 chemicals in DASS dataset in accordance with the analysis method described in the "Supporting document to the OECD guideline 497 on defined approaches for skin sensitisation" (OECD, 2021). The detailed analyses were summarized at Appendix 10. In summary, the EpiSensA dataset containing many pre/pro-haptens and lipophilic chemicals which can be considered as more challenging. However, EpiSensA showed good predictive performance, with only few limitations as described in the applicability domains identified.

An inter-laboratory ring study among the three participating laboratories was organized by the lead laboratory (Mizumachi, *et al.*, 2018). This study aimed at providing insight into the transferability, the within-laboratory reproducibility, and the between-laboratory reproducibility of the test method, by evaluating a set of 10 test chemicals with various sensitization potency and lipophilicity. In the context of this study, the EpiSensA method has been previously successfully transferred to two laboratories.

4. Scientific basis: biological and mechanistic relevance

OECD has reported an AOP of skin sensitization containing four key events in the initial step. KE2 suggests that inflammatory responses and gene expression of antioxidant response elements takes place in keratinocytes (OECD, 2012).

The gene expressions of *ATF3* and *IL-8* are involved in these inflammatory responses. It is known that ATF3 serves as a hub of a cellular adaptive-response network by negatively modulating the inflammatory cytokines and chemokines (Thompson *et al.*, 2009), and that the induction of *ATF3* is regulated by ATP and NFκB (Ohara *et al.*, 2010; Glichrist *et al.*, 2006). Moreover, haptens are known to induce ATP release from keratinocytes that express the ATP receptor called $P2X_7$ (Onami *et al.*, 2014), and the lack of $P2X_7$ results in resistance to contact hypersensitivity (Weber *et al.*, 2010). In addition, IL-8 serves as a potent chemotactic peptide for neutrophils (Leonard *et al.*, 1990); the latter being critically involved in both the sensitization and the elicitation phases of

contact hypersensitivity (Weber *et al.*, 2015). The induction of *IL-8* is regulated by the ATP-P2X$_7$ and p38MAPK pathways (Saito *et al.*, 2017; Hoffmann *et al.*, 2002).

It is known that two genes, the *GCLM* and the *DNAJB4*, are regulated by the nuclear factor E2-related factor 2 (Nrf2) and the antioxidant response element (ARE) (Saito *et al.*, 2017). GCLM serves as a key determinant of the biosynthesis of glutathione (GSH), that can regulate redox signaling and antioxidant response (Franklin *et al.*, 2008), and the induction of *GCLM* is regulated by both the Nrf2/ARE and the activator protein-1 (AP-1) pathways (Lu, 2013). Moreover, the molecular chaperone DNAJB4 suppresses protein misfolding by oxidative stress (Qiu *et al.*, 2006), and the induction of *DNAJB4* seems to be regulated by both the Nrf2/ARE pathway and the heat shock transcription factor-1/heat shock factor response element (HSF-1/HSE) pathway.

The possible mechanisms of the stimulus-specific regulation of the four aforementioned marker genes are described in Figure 4.

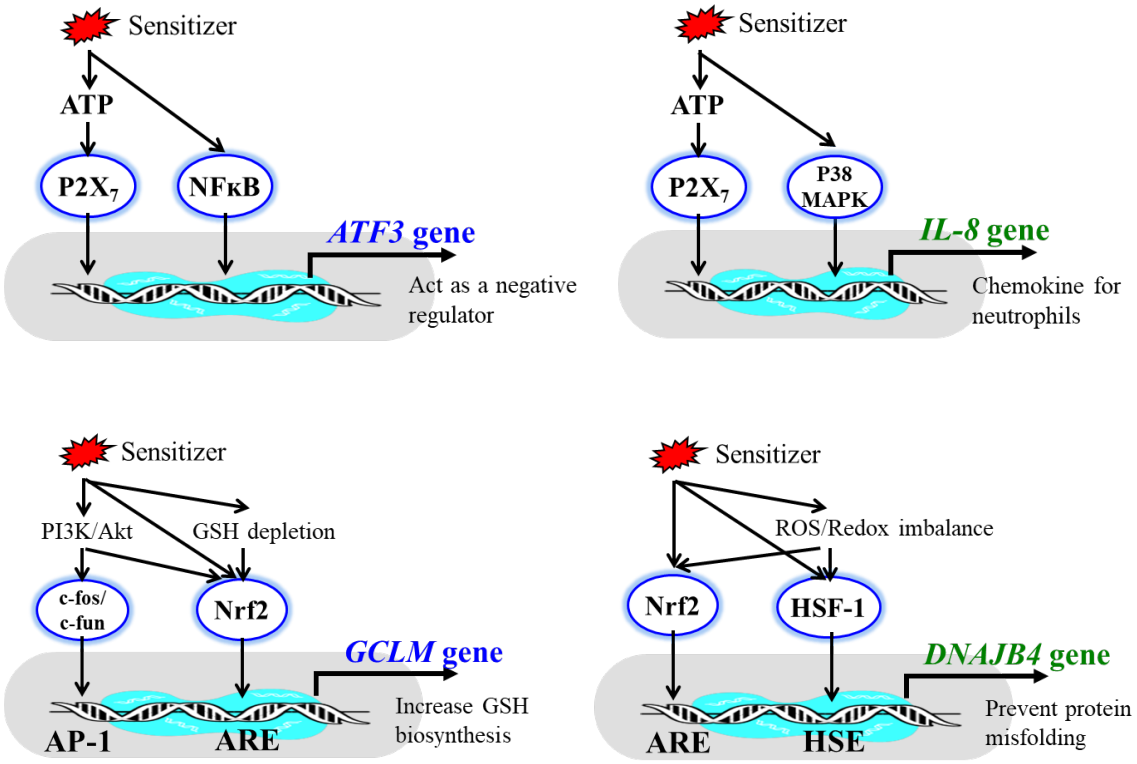


Figure 4. The possible mechanisms of the stimulus-specific regulation of four marker genes

5. SOP (Standard Operating Procedure)

The original version of the EpiSensA SOP is version 1.1 (see Appendix 7). In addition, an outline of the EpiSensA protocol is provided in Figure 5.

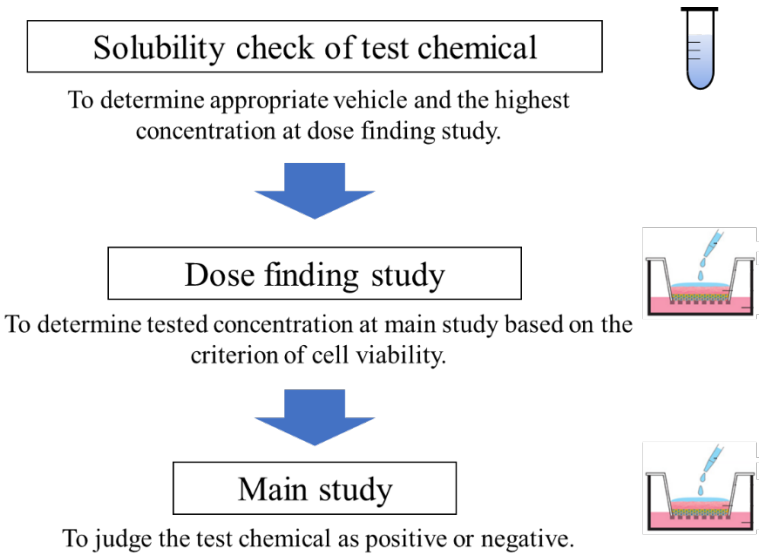


Figure 5. An outline of the EpiSensA protocol.

1) Solubility checks of the test chemicals

Each test chemical was dissolved in an appropriate vehicle, consisting of either AOO (acetone:olive oil at 4:1, v/v), distilled water (DW), or 50% ethanol in DW (50% EtOH). Specifically, the assay was performed by using ~~a chemical dissolved in~~ the vehicle that permitted the test chemical to be dissolved to the highest possible concentration.

2) Dose-finding study

The RhE model, namely, "LabCyte EPI-MODEL 24" (Japan Tissue Engineering Co. Ltd), was pre-cultured overnight at 37°C (5% CO_2) in 0.5 mL/well of culture medium provided by the manufacturer. For the undertaking of the dose-finding study, working solutions of each of the test chemicals were prepared as 4-fold serial dilutions ranging from the highest concentration to concentrations of 0.02% or below. An aliquot (5 µL) of the working solution was applied to the surface of each of tissue (1 tissue per group). One tissue for non-treated control and two tissues for the killed control (treated with 10 µL of 10% Triton X-100) were also prepared for the undertaking of the cell viability measurement. The tissues were incubated for 6 hours at 37°C under 5% CO_2. Cell

viability was measured through the lactate dehydrogenase (LDH) assay. The lowest test chemical concentration demonstrating less than 80% cell viability was used in the subsequent main study.

3) Main study

For the main study assessment of each test chemical, 2-fold serial dilutions were prepared so as to range from the lowest concentration resulting into less than 80% cell viability to the highest concentration allowing for more than 90% cell viability (basically, from 3 to 5 working solutions at 2-fold serial dilutions). When the viability was not less than 80% in the dose-finding study, the chemical was tested in the main study by preparing at least 3 working solutions at 2-fold serial dilutions from the highest soluble concentration. Aliquots (5 μL) of the working solutions of each chemical were applied to 3 tissues per group; the exposed tissues were then incubated for 6 h. One tissue for the non-treated control and 2 tissues for the killed controls were also prepared for the undertaking of the cell viability measurement. In addition, 3 tissues for each positive controls (at Transferability stage: 6.25% w/v bisphenol A diglycidyl ether, at Reliability stage: 0.78% w/v clotrimazole and 0.1% w/v 4-nitrobenzyl bromide) were also prepared.

Following the exposure to the test chemical, the tissue surface was rinsed three times with D-PBS, and the tissue was gently collected into a microtube. The tissues were then homogenized, and their total RNA extraction, cDNA synthesis, and real-time PCR were performed by using commercially-available reagents and kit. Cycle threshold (Ct) values of the four marker genes and of one endogenous control gene (*GAPDH*; encoding the housekeeping protein glyceraldehyde 3-phosphate dehydrogenase) were measured. Relative gene expression levels were calculated by using the $2^{-\Delta\Delta Ct}$ method and were expressed as fold-change normalized to the expression of the endogenous control gene (*GAPDH*).

The mean value (3 tissues per group) of the maximum fold-induction (Imax) was obtained by using the data from the concentrations allowing for more than or equal to 80% cell viability. When the Imax of at least one out of the four marker genes exceeded the respective cut-off values (*ATF3*: 15-fold, *GCLM*: 2-fold, *DNAJB4*: 2-fold, and *IL-8*: 4-fold), the chemical was judged as positive for EpiSensA.

6. Test Definition: Conclusions of the Validation Management Team

The EpiSensA test method addresses a biological mechanism – namely, the

inflammatory and cytoprotection-related gene expression at keratinocytes – that is considered to be a key step in the skin sensitization AOP. Information provided by the EpiSensA on a chemical's potential to induce the aforementioned keratinocyte response *in vitro* can complement the information generated by other non-animal approaches (*in silico*, *in chemico* or *in vitro*) designed to address the chemical/biological mechanisms preceding and following the keratinocyte response in the sequence of events leading to the induction of skin sensitization. This fact supports the use of EpiSensA as a mechanistically relevant element of a testing strategy for skin sensitization.

In conclusion, the VMT considers that the Test Method Definition is satisfied in this case.

Transferability

See also “Study plan of EpiSensA” (Appendix 1).

1. General aspects

The VMT considered that this study confirms that EpiSensA can be transferred to and performed at any laboratory where a standard real-time PCR equipment and technicians with the expertise to operate such equipment are available. All the equipment, labware and reagents needed to perform EpiSensA are commercially available.

Kao Corporation, as the lead laboratory in the development of EpiSensA, was responsible for ensuring that the test method described in the SOP was correctly understood and properly carried out by the participating laboratories, and did so by providing guidance, training and assistance to the laboratories participating in this validation study. The laboratory technicians at each participating laboratory, who received guidance from lead laboratory, have also undertaken training in their own laboratories.

The lead laboratory was also responsible for providing detailed plans for the technical transfer and the training based on its experience with the test method. The lead laboratory worked with the participating laboratories in order to implement the training testing and to ensure that a common understanding of and a compliance with the test acceptance criteria specified in the technology transfer plan. The test chemicals for each training phase were selected and supplied by the lead laboratory. The test chemicals were uncoded. The participating laboratories submitted their results promptly and accurately to the lead laboratory.

2. Training

Laboratory technicians and study directors from the three participating laboratories were provided with the technical transfer training for which both time information and guidance were also provided on the scientific rationale and principles of the test method, the EpiSensA SOP, the way to analyze the obtained results, and the way to apply the prediction model.

Date: 23rd May 2018, 9:30 ~ 17:30
Place: Hatano Research Institute Food and Drug Safety Center (FDSC)
Contents:
- Topical application of chemicals on the 3D skin model
- Tissue collection
- RNA extraction using TRIzol (Invitrogen) and the RNeasy Mini Kit (QIAGEN)

3. Transfer of the test method to participating laboratories

1) Goal

The purpose of the Transfer Phase was to assess the successful transfer of the protocol to the test facilities or operators that were inexperienced or poorly experienced with the use and functions of EpiSensA. The lead laboratory selected and distributed to each participating laboratory four uncoded test chemicals. The participating laboratories tested these four chemicals in accordance with the EpiSensA SOP. Proficiency was evaluated, determined and confirmed by analyzing the data from each laboratory before the VMT agreed to proceed with the Reliability Stage of the validation study.

2) Test chemicals

The four test chemicals chosen by the lead laboratory are shown in Table 3. Three out of the four chemicals (BADGE, clotrimazole and cetrimide) have been previously used in a ring study (Mizumachi *et al.*, 2018). BADGE was a previous positive control, and 4NBB was a candidate of the following (and now current) positive control. The selection of following positive control was concurrently performed with the technical transfer (see "**5. Modifications made to the original SOP**"). In addition, three out of these same four chemicals were known to be relatively easy to identify as either a sensitizer or a non-sensitizer, while the remaining one (cetrimide) was known to be relatively difficult to identify as either a sensitizer or a non-sensitizer.

BADGE, clotrimazole and cetrimide were assessed firstly through a dose-finding study, and secondly during the following main studies at the same condition that was decided based on the dose-finding study performed at each participating laboratory. In addition, 4NBB was tested twice at the indicated concentrations (0.1% and 0.2% w/v), as decided by the lead laboratory.

Table 3. Test chemicals used for technical transfer.

Chemical name	CAS No.	LLNA EC3(%)	GHS	LogKow	Vehicle
Bisphenol A diglycidyl ether (BADGE)	1675-54-3	1.5	1A	3.84	AOO
Clotrimazole	23593-75-1	4.8	1B	6.26	AOO
Cetrimide	57-09-0	-	Not classified	3.18	50% EtOH
0.1% and 0.2% w/v 4-Nitrobenzyl bromide (4NBB)	100-11-8	0.05	1A	2.70	AOO

3) Test conditions and success criteria

The participating laboratories conducted the tests in accordance with the EpiSensA SOP. BADGE (6.25% w/v) was used as a positive control during the main study at all technical transfer experiments. When all marker genes exceeded the respective cut-off values at 6.25% w/v BADGE, the results of main study were acceptable; if not, the results ought to be discarded. The experiments were performed once a week, and 1 to 3 chemicals were tested in each experiment. The number of test chemicals assessed in each experiment was decided by each operator.

When BADGE, clotrimazole and 4NBB were correctly judged as positive and cetrimide was correctly judged as negative in two main studies, the technical transfer was judged to complete successfully.

4. Results of technical transfer

1) LION

The transfer experiments started on the 13th June 2018. No major issues arose during the dose-finding study of the Transfer Phase. Five out of the eight main studies met the acceptance criteria with regard to the positive control. BADGE was tested in the main studies performed on the 20th of June 2018 and on the 8th of August 2018, and it was judged as positive in both studies consecutively. Clotrimazole was tested in the main studies performed on the 20th of June 2018 and on the 8th and 29th of August 2018. At the main study undertaken on the 20th of June 2018, a data loss occurred in one well as a result of an operational error. At the two subsequent main studies, clotrimazole was correctly judged as positive. Cetrimide was tested in the main studies performed on the

20th of June 2018 and on the 27th of June 2018, and it was judged as negative in both studies. Finally, 4NBB was tested in the main studies performed on the 8th and the 22nd of August 2018, and correctly judged as positive in both studies consecutively. Here, the Imax value of the *IL-8* as a result of the exposure to 4NBB exceeded the cut-off value (i.e., 4-fold) in the main study taking place on the 8th of August 2018, but not in the main study taking place on the 22nd of August 2018. It was confirmed that the fold induction of *IL-8* demonstrated a borderline induction when the 4NBB was tested in lead laboratory. In summary, all four test chemicals exhibited good reproducibility between the first and the second test, and were accurately predicted to be either sensitizers or non-sensitizers in agreement with the *in vivo* test results (Table 4).

Table 4. The results of the technical transfer experiments undertaken by LION (mean +/- S.D., n=3).

Chemical	GHS category	1st					2nd				
		Imax value				Judgment	Imax value				Judgment
		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)	
BADGE	Cat. 1A	**50.1 ± 6.7**	**6.7 ± 0.7**	**15.2 ± 2.8**	**12.1 ± 1.0**	Positive	**68.4 ± 6.2**	**7.1 ± 1.0**	**20.2 ± 0.9**	**15.9 ± 0.4**	Positive
Clotrimazole	Cat. 1B	**925.5 ± 437.8**	1.1 ± 0.1	**33.7 ± 13.4**	**140.7 ± 49.0**	Positive	**823.2 ± 189.6**	1.8 ± 0.2	**54.3 ± 12.2**	**146.3 ± 28.6**	Positive
Cetrimide	Not classified	12.0 ± 11.9	1.0 ± 0.2	1.1 ± 0.0	1.5 ± 0.5	Negative	11.2 ± 3.2	0.9 ± 0.0	1.1 ± 0.1	2.0 ± 0.2	Negative
4NBB	Cat. 1A	**170.5 ± 2.7**	**12.5 ± 2.0**	**20.9 ± 1.0**	**4.2 ± 0.6**	Positive	**47.4 ± 6.5**	**13.2 ± 0.7**	**24.1 ± 3.2**	1.9 ± 0.4	Positive

2) KOSÉ

The transfer experiments started on the 6th of June 2018. No major issues arose during the dose-finding study of the Transfer Phase. Four out of the five main studies met the acceptance criteria with regard to the positive control. BADGE, clotrimazole and cetrimide were tested in the main studies performed on the 13th and 27th of June 2018. Both BADGE and clotrimazole were correctly judged as positive in both studies consecutively. In addition, cetrimide was correctly judged as negative in both studies. 4NBB was tested in the main studies performed on the 25th of July 2018 and on the 8th of August 2018, and correctly judged as positive in both studies. In summary, all four test chemicals exhibited good reproducibility between the first and the second test, and were accurately predicted to be either sensitizers or non-sensitizers in agreement with the *in vivo* test results in two main studies (Table 5).

Table 5. The results of the technical transfer experiments undertaken by KOSÉ (mean +/- S.D., n=3).

Chemical	GHS category	1st					2nd				
		Imax value				Judgment	Imax value				Judgment
		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)	
BADGE	Cat. 1A	**21.8 ± 3.1**	**2.8 ± 0.6**	**6.0 ± 1.5**	**20.9 ± 5.4**	Positive	**45.1 ± 7.7**	**4.7 ± 0.6**	**14.0 ± 1.4**	**46.8 ± 4.3**	Positive
Clotrimazole	Cat. 1B	**896.7 ± 505.1**	1.5 ± 0.2	**28.2 ± 15.8**	**223.6 ± 121.3**	Positive	**1059.0 ± 152.3**	1.8 ± 0.0	**43.9 ± 3.6**	**299.4 ± 45.4**	Positive
Cetrimide	Not classified	2.4 ± 1.7	0.8 ± 0.2	0.9 ± 0.1	1.4 ± 0.5	Negative	9.5 ± 4.6	1.6 ± 0.1	1.2 ± 0.1	3.5 ± 1.0	Negative
4NBB	Cat. 1A	**120.5 ± 73.4**	**8.0 ± 7.5**	**17.2 ± 12.5**	**10.5 ± 7.6**	Positive	**24.1 ± 2.1**	**6.0 ± 0.7**	**6.3 ± 0.6**	**5.1 ± 0.6**	Positive

3) FDSC

There were two operators at the FDSC and both of them performed the technical transfer experiments individually. As far as the first operator is concerned, the transfer experiments started on the 13th of June 2018, and no major issues arose during the dose-finding study of the Transfer Phase. Four out of the five performed main studies met the acceptance criteria regarding the positive control. BADGE and clotrimazole were tested in the main studies performed on the 13th of June 2018 and on the 18th of July 2018. Both BADGE and clotrimazole were correctly judged as positive in both studies. Cetrimide was tested in the main studies performed on the 13th of June 2018, on the 18th of July 2018, as well as on the 1st and 22nd of August 2018. Cetrimide was correctly judged as negative on the 13th of June 2018, but it was wrongly judged as positive on the 18th of July 2018 and on the 1st of August 2018 because the *ATF3* and/or the *IL-8* expressions slightly exceeded the respective cut-off values. It was later confirmed that the change fold of the *ATF3* and *IL-8* expression was increased near to the respective cut-off values when cetrimide was tested in the lead laboratory. At the subsequent main study that took place on the 22nd of August 2018, cetrimide was correctly judged as negative. 4NBB was tested in the main studies performed on the 18th of July 2018 and on the 1st of August 2018, and correctly judged as positive in both studies consecutively. In summary, all four test chemicals were accurately predicted to be either sensitizers or non-sensitizers in agreement with the *in vivo* test results in two main studies (Table 6).

Table 6. The results of the technical transfer experiments undertaken by the first operator at FDSC (mean +/- S.D., n=3).

Chemical	GHS category	1st					2nd				
		Imax value				Judgment	Imax value				Judgment
		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)	
BADGE	Cat. 1A	**65.9 ± 11.6**	**6.8 ± 0.8**	**26.5 ± 0.9**	**17.6 ± 3.3**	Positive	**54.0 ± 7.6**	**3.0 ± 0.3**	**21.0 ± 5.8**	**25.6 ± 3.1**	Positive
Clotrimazole	Cat. 1B	**597.3 ± 222.2**	1.6 ± 0.5	**33.1 ± 11.0**	**105.0 ± 13.6**	Positive	**606.4 ± 171.9**	1.9 ± 0.2	**58.4 ± 6.8**	**103.9 ± 23.7**	Positive
Cetrimide	Not classified	11.8 ± 4.0	1.2 ± 0.1	1.3 ± 0.1	3.1 ± 0.5	Negative	12.8 ± 2.1	1.0 ± 0.1	1.1 ± 0.1	3.2 ± 0.2	Negative
4NBB	Cat. 1A	**77.6 ± 22.7**	**4.5 ± 0.1**	**13.1 ± 1.3**	2.5 ± 0.6	Positive	**129.3 ± 8.9**	**7.6 ± 1.2**	**12.2 ± 2.4**	3.8 ± 0.6	Positive

Regarding the second operator, the transfer experiments started on the 13th of June 2018. No major issues arose during the dose-finding study of the Transfer Phase. Eight out of the ten main studies met the acceptance criteria set for the positive controls. BADGE and clotrimazole were tested in the main studies performed on the 20th of June 2018 and on the 25th of July 2018. Both BADGE and clotrimazole were correctly judged as positive in both studies. Cetrimide was tested in the main studies performed on the 13th of June 2018, on the 18th of July 2018, as well as on the 1st and 22nd of August 2018. Cetrimide was tested seven times in total and was correctly judged as negative in the main studies undertaken on the 25th of July 2018, as well as on the 12th and the 19th of September 2018. At the remaining four main studies, cetrimide was falsely judged as positive due to the expression of *ATF3* and/or *IL-8* being slightly exceeding the respective cut-off values. It was confirmed that the fold induction of the *ATF3* and *IL-8* expressions increased closely to the respective cut-off values when cetrimide was tested in lead laboratory. Eventually, cetrimide was consecutively judged as negative in the main studies undertaken on the 12th and the 19th of September 2018. Finally, 4NBB was tested in the main studies performed on the 25th of July 2018 and on the 8th of August 2018, and correctly judged as positive in both studies. In summary, all four test chemicals were accurately predicted to be either sensitizers or non-sensitizers, in agreement with the *in vivo* test results in two main studies (Table 7).

Table 7. The results of the technical transfer experiments undertaken by the second operator at FDSC (mean +/- S.D., n=3).

Chemical	GHS category	1st					2nd				
		Imax value				Judgment	Imax value				Judgment
		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)	
BADGE	Cat. 1A	**24.4 ± 2.9**	**5.9 ± 0.4**	**24.0 ± 2.3**	**14.1 ± 1.5**	Positive	**24.9 ± 6.1**	**2.9 ± 0.2**	**14.8 ± 3.1**	**9.7 ± 1.7**	Positive
Clotrimazole	Cat. 1B	**466.6 ± 109.4**	1.9 ± 0.0	**58.9 ± 20.5**	**144.8 ± 40.3**	Positive	**388.1 ± 91.8**	1.3 ± 0.3	**46.3 ± 10.6**	**111.3 ± 31.5**	Positive
Cetrimide	Not classified	3.0 ± 1.1	1.0 ± 0.1	1.1 ± 0.1	2.1 ± 0.7	Negative	6.7 ± 5.4	0.9 ± 0.0	1.1 ± 0.2	2.3 ± 0.7	Negative
4NBB	Cat. 1A	**57.8 ± 26.0**	**7.1 ± 4.7**	**17.2 ± 11.2**	2.1 ± 1.4	Positive	**76.8 ± 42.9**	**6.5 ± 4.9**	**15.3 ± 5.1**	3.7 ± 0.7	Positive

5. Modifications made to the original SOP

The original SOP for this validation study was modified by the lead laboratory. Based on the comments at the joint meeting of the VMT (that took place from the 3rd to the 5th of July 2018), the SOP was modified in order to remove ambiguities and to resolve minor technical issues. These modifications were thus included in version 2.1 of the SOP. The most significant revisions were as follows:

- Positive control was changed from 6.25% w/v bisphenol A diglycidyl ether (BADGE) to 0.78 w/v% clotrimazole and 0.10% w/v 4-nitrobenzylbromide (4NBB) because of the poor stability and the inaccurate concentration of the commercially available BADGE. As the acceptance criteria, the Imax values of all marker genes were required to exceed the respective cut-off values at 0.1 or 0.2% w/v of 4NBB with good reproducibility (exceed the cut-off value in all experiments) at all participating laboratories. If that was not the case, then more than one positive control would be set so as all marker genes exceeded the respective cut-off values for at least one positive control. As a result, *GCLM* and *DNAJB4* exceeded their respective cut-off value in all experiments at all laboratories. However, *ATF3* and *IL-8* did not exceed the respective cut-off values at all experiments, and the aforementioned criteria was not met. Therefore, 4NBB and clotrimazole were assigned as positive control chemicals as for them both *ATF3* and *IL-8* exceeded their cut-off values with good reproducibility when exposed to clotrimazole. The concentration of 4NBB was decided to be 0.1% w/v due to its ability to significantly induce *GCLM* and *DNAJB4* at this concentration rather than at that of 0.2% w/v. Furthermore, the concentration of clotrimazole was decided to be that of 0.78% w/v, because of its ability to cause a multifold induction of *ATF3* and *IL-8* similar to that caused by 6.25% w/v BADGE.

- The acceptance criteria for the mean viability of the vehicle control has changed from 80% to 90%. This is because the gene expressions of *ATF3* and *IL-8* seem to slightly increase along with the decrease in cell viability, suggesting that the background level might lead to an underestimation of the sensitizing potential of the test chemicals (Figure 6). The new criterion derived from the statistical analysis on the cell viability distribution of the vehicle control at a total of 96 experiments undertaken by the lead laboratory (Figure 7).

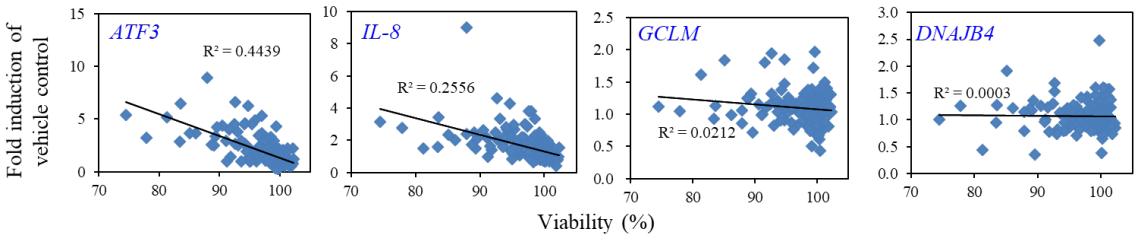


Figure 6. Relationship between the fold induction of four marker genes of the AOO control (*vs*. non-treated) and the cell viability (n = 183).

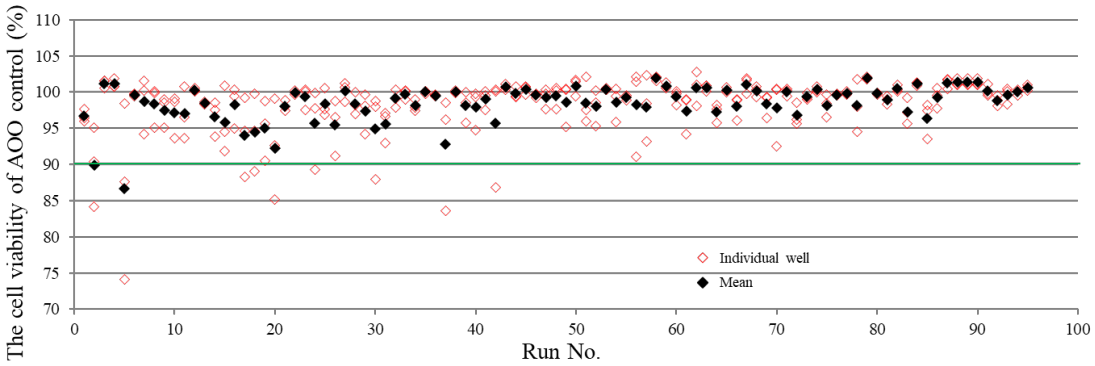


Figure 7. The cell viability distribution of the vehicle control throughout a total of 96 experiments undertaken by the lead laboratory.

- The acceptance criterion for the *GAPDH* Ct value of the test chemicals has been added. When the mean *GAPDH* Ct value of each tested concentration is within the mean *GAPDH* Ct value of the corresponding vehicle control +/- 1, the result at that specific concentration can be considered as acceptable. The criterion was based on a total of 1,337 historical datapoints archived by the lead laboratory.

6. Conclusions of the VMT

The transfer of the EpiSensA from the lead laboratory to the three participating laboratories was successfully completed and to the satisfaction of both the lead laboratory and the VMT.

Three sensitizers, that were known to be relatively easy to identify as such, were correctly judged as positive in all of the main studies undertaken by all laboratories, with good reproducibility. Cetrimide was correctly judged as negative in all main studies undertaken by KOSÉ and LION, but it was falsely judged as positive in several of the main studies performed at FDSC. However, as described above, cetrimide is known to be relatively difficult to identify as a non-sensitizer, since the fold induction of the *ATF3* and *IL-8* expressions increase near to the respective cut-off values when tested at the lead laboratory. It is noteworthy that cetrimide is used as a surfactant or a fungicide, and it does have a skin irritation potential. The increases in the gene expression of *ATF3* and *IL-8* are also observed along with a decrease in cell viability when other surfactants (such as sodium lauryl sulfate and benzalkonium chloride) are tested on EpiSensA. Therefore, the observed increases of the *ATF3* and the *IL-8* expression might not necessarily be skin sensitization-specific reactions.

Within-Laboratory Reproducibility

The within-laboratory reproducibility study (Phase I study) was divided into three parts, hereby abbreviated as Phase I-A, I-B and I-C. The SOP version 2.1 was used during the undertaking of Phase I-A at first, but retests were performed in accordance with the SOP version 2.2 if necessary (see "1-4. Revision of the SOP"). During the Phases I-B and I-C, the SOP versions 2.2 and 2.3 were in use, respectively.

In order to minimize bias in the generation of the results for the assessment of the WLR, the participating laboratories were provided with 3 different vials for each chemical (with different codes); one for each experiment. For this reason, the laboratories had to perform a separate solubility check, vehicle determination and dose-finding study, prior to the performance of the main study; these were done on three separate occasions for each chemical.

The WLR of EpiSensA was assessed on the concordance of classification, the assignment of a "positive" (P) versus a "negative" (N) label, amongst three independent experiments. Additionally, descriptive analyses were performed in order to evaluate the reproducibility of the three respective Imax values for the induction of expression of the four marker genes.

1. Phase I-A

1-1. LION

a. Reproducibility (concordance in prediction)

Three sensitizers and two non-sensitizers were tested three times, respectively. Prior to retesting, the reproducibility in terms of their classification as P versus N, confirmed the same prediction in all three independent experiments and was obtained for 5 out of the 5 chemicals tested, resulting in a WLR of 100%. However, benzisothiazolinone (chemical #3) was retested in accordance with the SOP version 2.2 at the 1st, 2nd and 3rd experiment, due to the fact that the SOP version 2.1 had some ambiguity and inappropriate cases in it. Finally, the same prediction in the three independent experiments was obtained for 4 out of the 5 chemicals tested, resulting in a WLR of 80%. The chemical for which the classification was not concordant across all experiments was benzisothiazolinone (chemical #3) (Table 8).

Table 8. The results of Phase I-A, as conducted by LION.

No.	Chemical	Exp.1			Exp.2			Exp.3			Agreement between experiment
		Vehicle	Judge	Retest	Vehicle	Judge	Retest	Vehicle	Judge	Retest	
1	Glyoxal 40% solution in water	DW	P	-	DW	P	-	DW	P	-	Yes
2	Lauryl gallate	AOO	N	-	AOO	N	-	AOO	N	-	Yes
3	Benzisothiazolinone	AOO	N	P	AOO	N	N	AOO	N	N	No
4	Diethylphthalate	AOO	P	-	AOO	P	-	AOO	P	-	Yes
5	Sodium lauryl sulfate	DW	P	-	DW	P	-	DW	P	-	Yes

b. Reproducibility of the Imax value

Figure 8 presents the distribution of the Imax values (maximum fold-induction with more than or equal to 80% cell viability) as assessed through three independent experiments for each of the 5 chemicals tested. As far as the Imax values of *ATF3* are concerned, the values of chemical #1, 2, 3, and 5 showed concordant results in terms of whether they exceed the cut-off value or not, but the values of chemical #4 did not. Therefore, concordant results across the three independent experiments were demonstrated at 4/5 chemicals. Likewise, concordant results were demonstrated at 4/5, 4/5, and 5/5 chemicals in *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, notwithstanding that different RhE batches have been used, the Imax values for each of the marker genes had good consistency; in fact, a total of 17/20 demonstrated concordance across the three independent experiments performed.

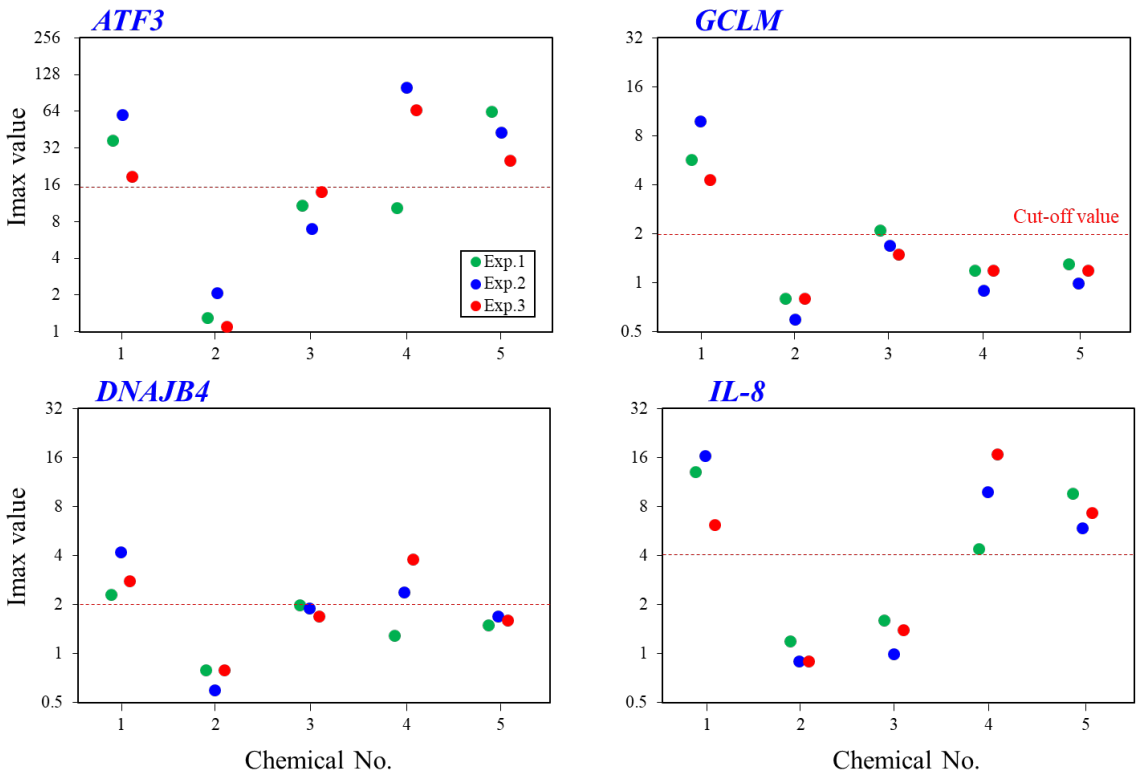


Figure 8. Distribution of the mean Imax values (n=3) for chemicals #1-5, as assessed by three individual experiments at LION. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

1-2. KOSÉ

a. Reproducibility (concordance in prediction)

Prior to retesting, the reproducibility in terms of their classification as P versus N, confirmed the same prediction in all three independent experiments and was obtained for 4 out of the 5 chemicals tested, resulting in a WLR of 80%. However, sodium lauryl sulfate (chemical #5) at the 2nd experiment was retested in accordance with the SOP version 2.2, due to the SOP version 2.1 having some ambiguity and inappropriate cases' mentioning. Finally, the same prediction was obtained by the three independent experiments for 5 out of the 5 chemicals tested, resulting in a WLR of 100% (Table 9).

Table 9. The results of Phase I-A, as conducted by KOSÉ.

No.	Chemical	Exp.1			Exp.2			Exp.3			Agreement between experiment
		Vehicle	Judge	Retest	Vehicle	Judge	Retest	Vehicle	Judge	Retest	
1	Glyoxal 40% solution in water	DW	P	-	DW	P	-	DW	P	-	Yes
2	Lauryl gallate	AOO	N	-	AOO	N	-	AOO	N	-	Yes
3	Benzisothiazolinone	AOO	P	-	AOO	P	-	AOO	P	-	Yes
4	Diethylphthalate	AOO	P	-	AOO	P	-	AOO	P	-	Yes
5	Sodium lauryl sulfate	50% EtOH	P	-	50% EtOH	N	P	50% EtOH	P	-	Yes

b. Reproducibility of the Imax value

Figure 9 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. As far as the Imax values of *ATF3* are concerned, the values of chemical #2, 4, and 5 showed concordant results in terms of whether they exceed the cut-off value or not, but the values of chemical #1 and 3 did not. Therefore, concordant results across the three independent experiments were demonstrated at 3/5 chemicals. Likewise, concordant results were demonstrated at 4/5, 4/5, and 4/5 chemicals in *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, notwithstanding that different RhE batches have been used, the Imax values for each of the marker genes had good consistency; in fact, a total of 15/20 demonstrated concordance across the three independent experiments performed.

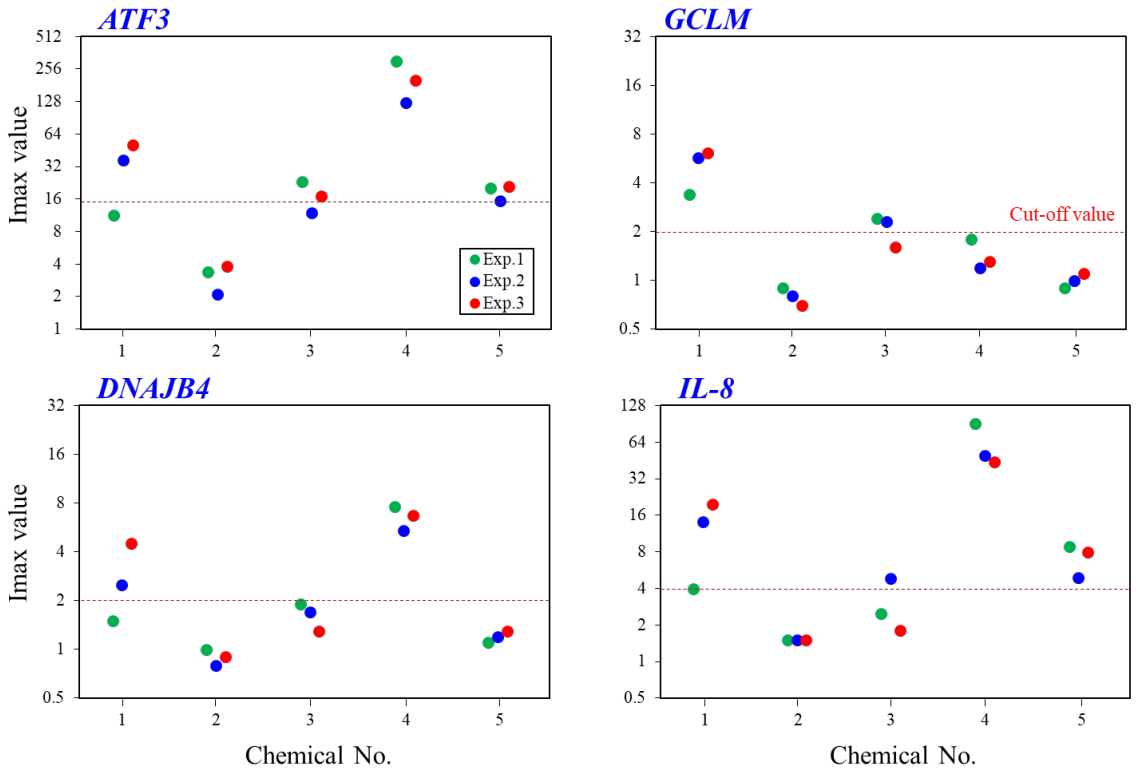


Figure 9. Distribution of the mean Imax values (n=3) for chemicals #1-5, as assessed by three individual experiments at KOSÉ. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

1-3. FDSC

a. Reproducibility (concordance in prediction)

Prior to retesting, the reproducibility in terms of their classification as P versus N, confirmed the same prediction in all three independent experiments and was obtained for 3 out of the 5 chemicals tested, resulting in a WLR of 60%. However, benzisothiazolinone (chemical #3) at 3rd experiment was retested in accordance with the SOP version 2.2 instead of version 2.1. Finally, the same prediction was obtained by the three independent experiments for 4 out of 5 of the chemicals tested, resulting in a WLR of 80%. The chemical for which the classification was not concordant across all experiments was lauryl gallate (chemical #2) (Table 10).

Table 10. The results of Phase I-A, as conducted by FDSC.

No.	Chemical	Exp.1			Exp.2			Exp.3			Agreement between experiment
		Vehicle	Judge	Retest	Vehicle	Judge	Retest	Vehicle	Judge	Retest	
1	Glyoxal 40% solution in water	DW	P	-	DW	P	-	DW	P	-	Yes
2	Lauryl gallate	AOO	N	-	AOO	P	-	AOO	N	-	No
3	Benzisothiazolinone	AOO	P	-	AOO	P	-	AOO	N	P	Yes
4	Diethylphthalate	AOO	P	-	AOO	P	-	AOO	P	-	Yes
5	Sodium lauryl sulfate	50% EtOH	P	-	50% EtOH	P	-	50% EtOH	P	-	Yes

b. Reproducibility of the Imax value

Figure 10 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. The concordant results were demonstrated at 3/5, 4/5, 3/5, and 4/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, the Imax values for each of the marker genes had good consistency; a total of 14/20 demonstrated concordance across the three independent experiments performed.

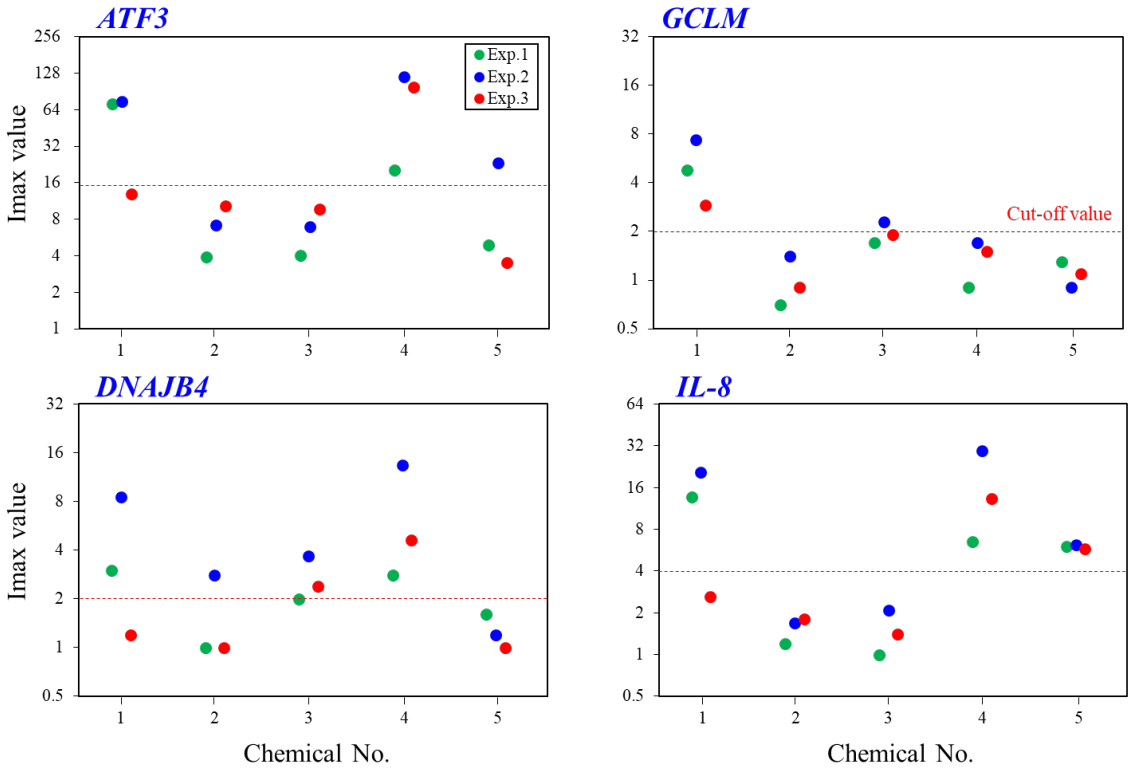


Figure 10. Distribution of the mean Imax values (n=3) for chemicals #1-5, as assessed by three individual experiments at FDSC. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

1-4. Revision of the SOP and other consideration

The following revisions were proposed, discussed and approved at a web meeting of the VMT (that took place on the 4th of September 2019) after the end of the Phase I-A study. According to the modified SOP, the participating laboratories would retest the test chemicals of Phase I-A if necessary. The following modifications were thus included in the SOP version 2.2:

- The acceptance criteria for the vehicle control has been changed again; each of the cell viability of at least two individual epidermises of the vehicle control should be equal to or greater than 95%. In addition, the acceptance criteria for the mean cell viability of the vehicle control was removed. During the Phase I-A study, there have been several cases in which one out of the three epidermises of the vehicle control have demonstrated a markedly low viability despite the fact that the mean cell viability of the three epidermises of the vehicle control have met the acceptance criterion of being ≥90% (Figure 11), and it was confirmed that a well with <90% cell viability at the vehicle control might cause an underestimation of test chemicals' sensitizing potential. During the analysis of the lead laboratory data regarding the *ATF3* induction as a result of the previous positive control (6.25% w/v BADGE), when the data included a well with a <90% cell viability at the vehicle control, this set of data was removed; as a result, the fold inductions of *ATF3* in the case of BADGE were allowed to increase (from blue diamonds to red diamonds in Figure 12) and were subsequently become similar to those observed in the experiments where all wells were exhibiting a >90% cell viability at the case of their vehicle control (by comparing the red and the green diamonds in Figure 12). This approach has suggested that the underestimation could have been prevented by removing the data with a <90% cell viability at vehicle control. The new acceptance criterion (requiring a ≥95% cell viability in each individual vehicle control well) was set based on the statistical analysis of the cell viability distribution of the vehicle control obtained by lead laboratory (red diamonds in Figure 7).

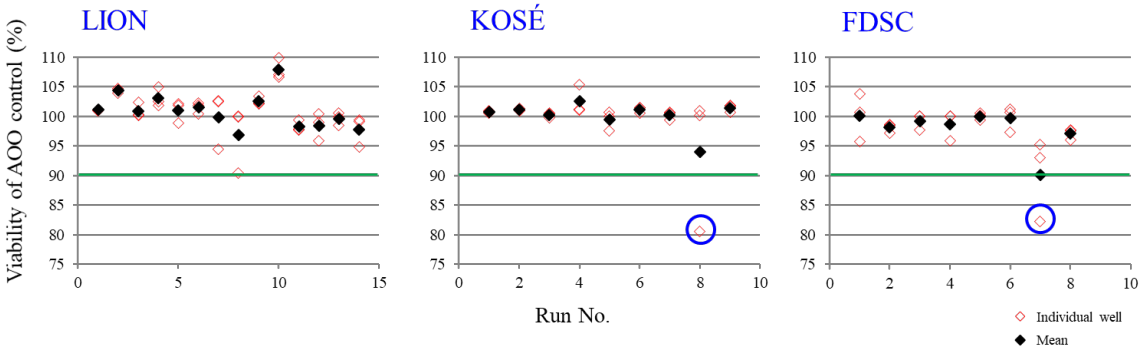


Figure 11. Cases in which 1 out of the 3 epidermises of the vehicle control showed a markedly low cell viability during Phase I-A.

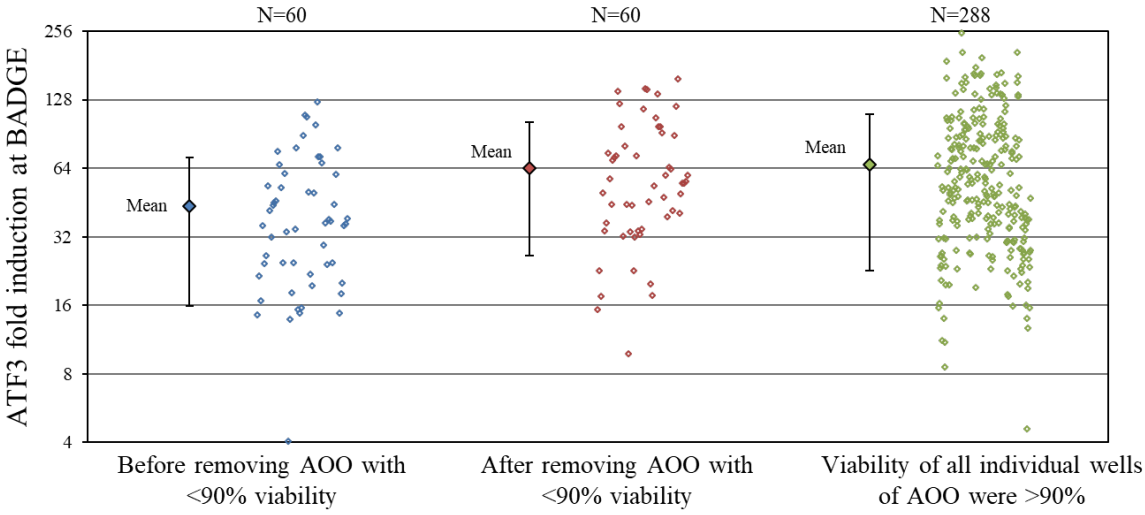
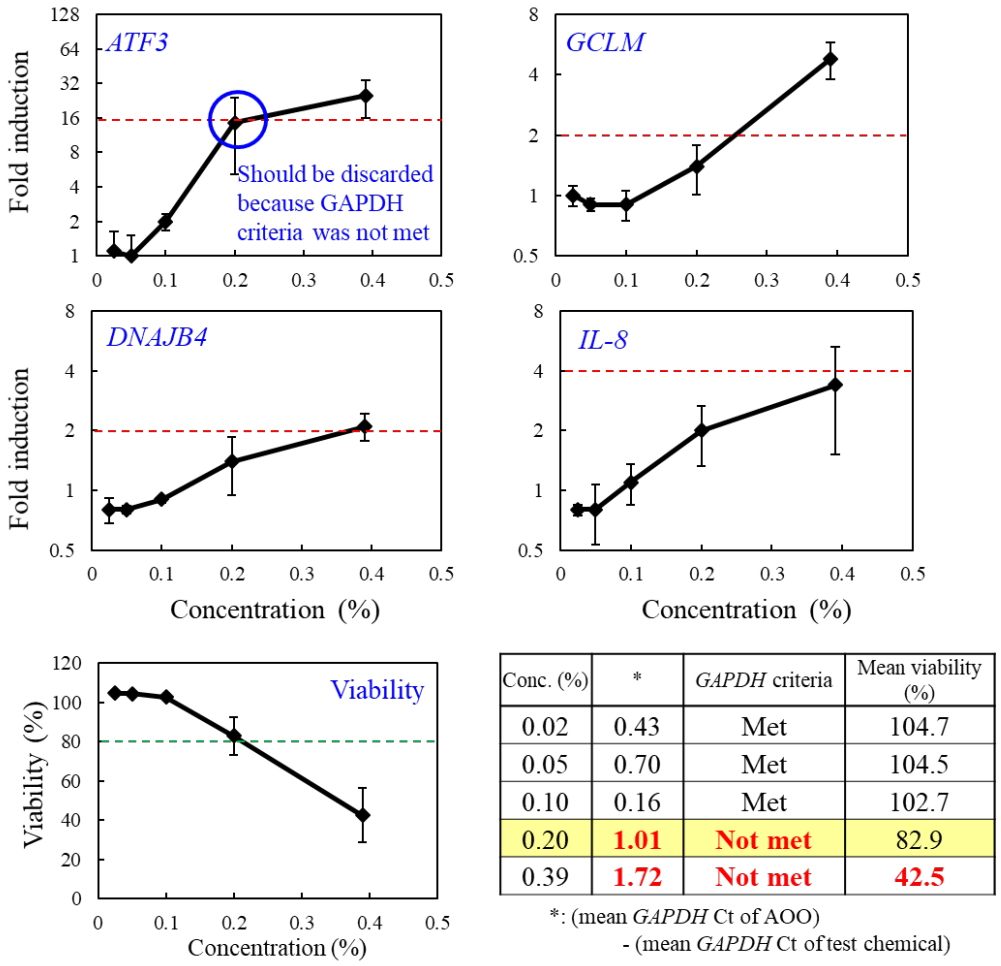


Figure 12. The effect of the vehicle control (AOO) with or without the inclusion of a well demonstrating a cell viability of <90% on the *ATF3* fold induction as a result of exposure to 6.25% w/v BADGE.

- The retesting criterion concerning a chemical with a steep dose-response curve was clarified. When a test chemical shows a steep or a greatly fluctuating dose-response curve for cell viability and/or gene induction, and the fold change of the gene expression exceeds the cut-off value just at the lowest concentration with less than 80% mean cell viability, then this test chemical should be retested with a narrower dose-response analysis by using a lower dilution factor (e.g., √2 (=1.41) fold dilution), in order to determine whether the induction has occurred at cytotoxic levels (80 to

95% mean cell viability). A 95% of cell viability is set as the upper limit as a result of the analysis required for the validity of the cell viability produced by the vehicle control (see "**5. Modifications made to the original SOP**" at the Transferability section).

- The retesting criterion regarding the *GAPDH* Ct value was modified. During the Phase I-A study, there is a possibility that the *GAPDH* Ct value is very slightly out of the range of the acceptance criteria at a test concentration generating a positive increase of the marker gene fold change and an acceptable cell viability (Figure 13). Therefore, the criterion was modified as follows: if the *GAPDH* Ct value does not meet the acceptance criteria at the highest tested concentration with an equal to or greater than 80% mean cell viability and – at the same time – the fold change in gene expression does not exceed the respective cut-off values at the lower concentrations, then the test chemical should be retested.



Conc. (%)	*	*GAPDH* criteria	Mean viability (%)
0.02	0.43	Met	104.7
0.05	0.70	Met	104.5
0.10	0.16	Met	102.7
0.20	**1.01**	**Not met**	82.9
0.39	**1.72**	**Not met**	**42.5**

*: (mean *GAPDH* Ct of AOO)
- (mean *GAPDH* Ct of test chemical)

Figure 13. The case in which the *GAPDH* Ct value was very slightly out of the range of the acceptance criteria at a test concentration exerting a positive increase of the marker gene expression and an acceptable cell viability (n=3, bars: S.D.).

2. Phase I-B

2-1. LION

a. Reproducibility (concordance in prediction)

Two sensitizers and three non-sensitizers were tested three times each. The same prediction in all three independent experiments was obtained for 5 out of the 5 chemicals tested, resulting in a WLR of 100% (Table 11).

Table 11. The results of Phase I-B, as conducted by LION.

No.	Chemical	Exp.1		Exp.2		Exp.3		Agreement between experiment
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
6	Hexane	AOO	N	AOO	N	AOO	N	Yes
7	Dextran	DW	N	DW	N	DW	N	Yes
8	Tween80	AOO	N	AOO	N	AOO	N	Yes
9	Ethyl acrylate	AOO	P	AOO	P	AOO	P	Yes
10	2,4-Dinitrochlorobenzene	AOO	P	AOO	P	AOO	P	Yes

b. Reproducibility of Imax value

Figure 14 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. The concordant results were demonstrated at 5/5, 5/5, 5/5, and 4/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, the Imax values for each of the marker genes had good consistency; a total of 19/20 demonstrated concordance across the three independent experiments performed.

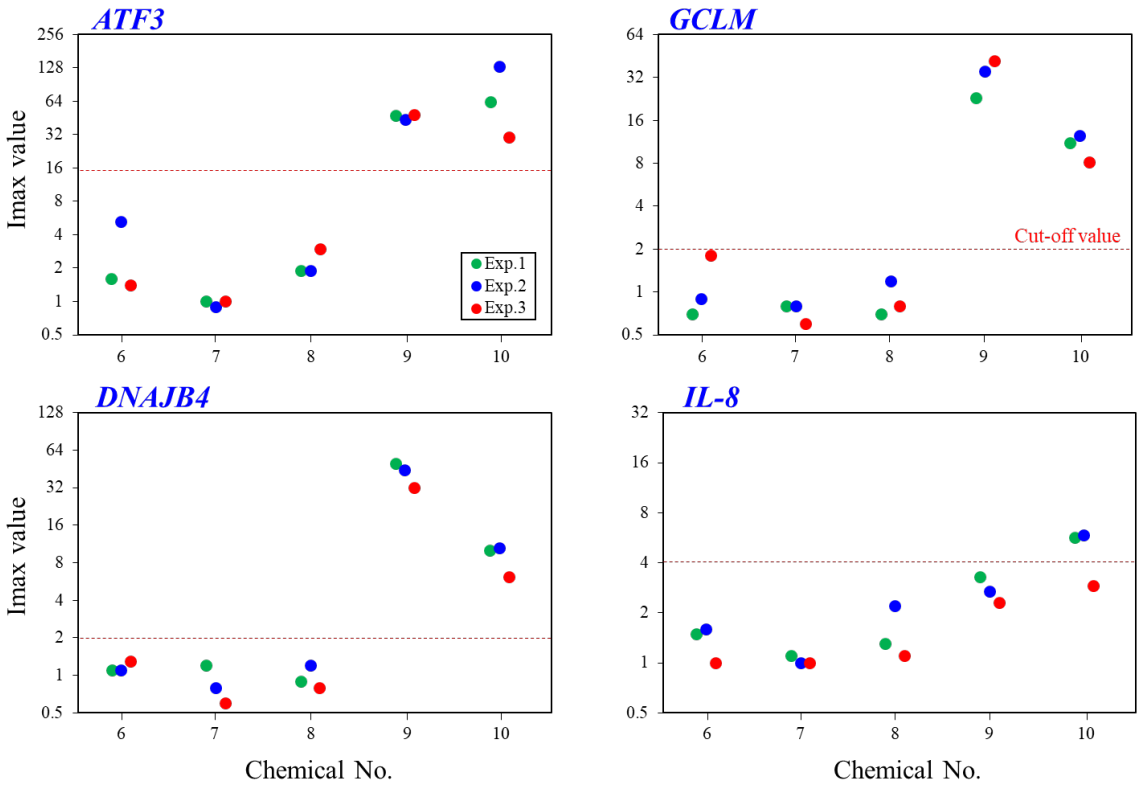


Figure 14. Distribution of the mean Imax values (n=3) for chemicals #6-10, as assessed by three individual experiments at LION. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

2-2. KOSÉ

a. Reproducibility (concordance in prediction)

The same prediction in all three independent experiments was obtained for 5 out of the 5 chemicals tested, resulting in a WLR of 100% (Table 12).

Table 12. The results of Phase I-B, as conducted by KOSÉ.

No.	Chemical	Exp.1		Exp.2		Exp.3		Agreement between experiment
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
6	Hexane	AOO	N	AOO	N	AOO	N	Yes
7	Dextran	DW	N	DW	N	DW	N	Yes
8	Tween80	AOO	N	AOO	N	AOO	N	Yes
9	Ethyl acrylate	AOO	P	AOO	P	AOO	P	Yes
10	2,4-Dinitrochlorobenzene	AOO	P	AOO	P	AOO	P	Yes

b. Reproducibility of the Imax value

Figure 15 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. The concordant results were demonstrated at 5/5, 5/5, 5/5, and 4/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, the Imax values for each of the marker genes had good consistency; a total of 19/20 demonstrated concordance across the three independent experiments performed.

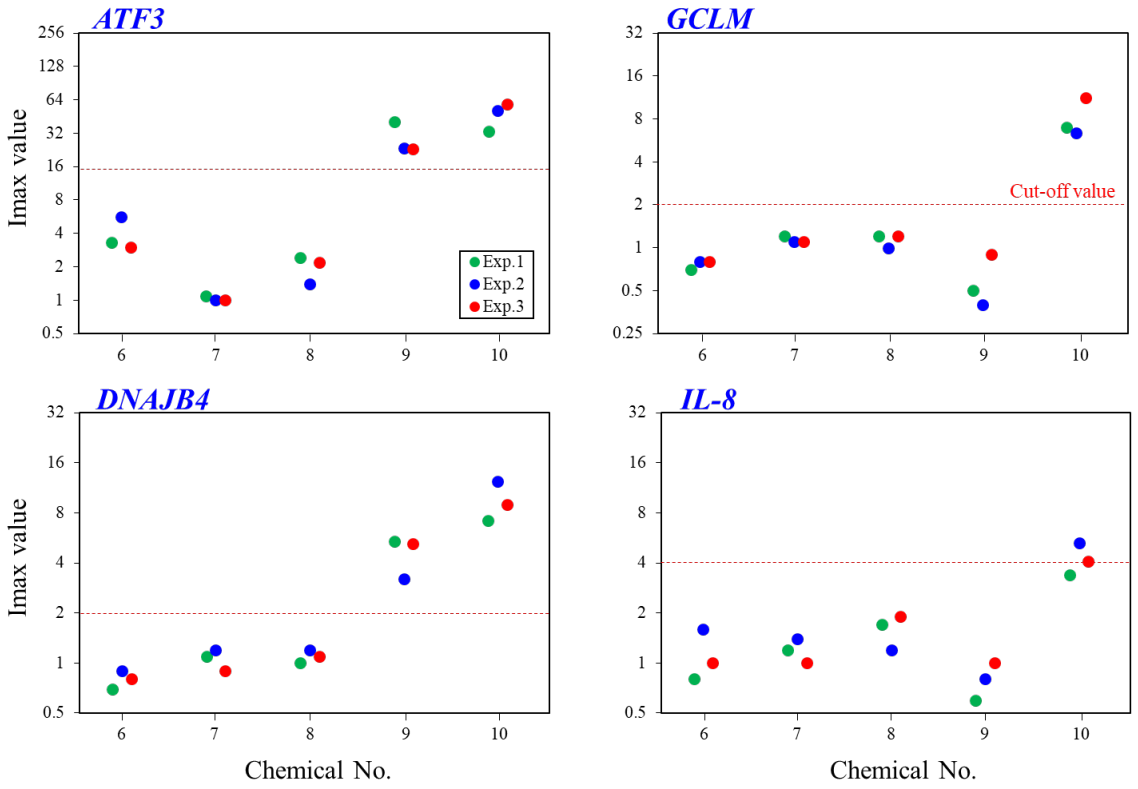


Figure 15. Distribution of the mean Imax values (n=3) for chemicals #6-10, as assessed by three individual experiments at KOSÉ. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

2-3. FDSC

a. Reproducibility (concordance in prediction)

The same prediction in all three independent experiments was obtained for 5 out of the 5 chemicals tested, resulting in a WLR of 100% (Table 13).

Table 13. The results of Phase I-B, as conducted by FDSC.

No.	Chemical	Exp.1		Exp.2		Exp.3		Agreement between experiment
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
6	Hexane	AOO	N	AOO	N	AOO	N	Yes
7	Dextran	DW	N	DW	N	DW	N	Yes
8	Tween80	AOO	N	AOO	N	AOO	N	Yes
9	Ethyl acrylate	AOO	P	AOO	P	AOO	P	Yes
10	2,4-Dinitrochlorobenzene	AOO	P	AOO	P	AOO	P	Yes

b. Reproducibility of the Imax value

Figure 16 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. The concordant results were demonstrated at 4/5, 5/5, 5/5, and 5/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, the Imax values for each of the marker genes had good consistency; a total of 19/20 demonstrated concordance across the three independent experiments performed.

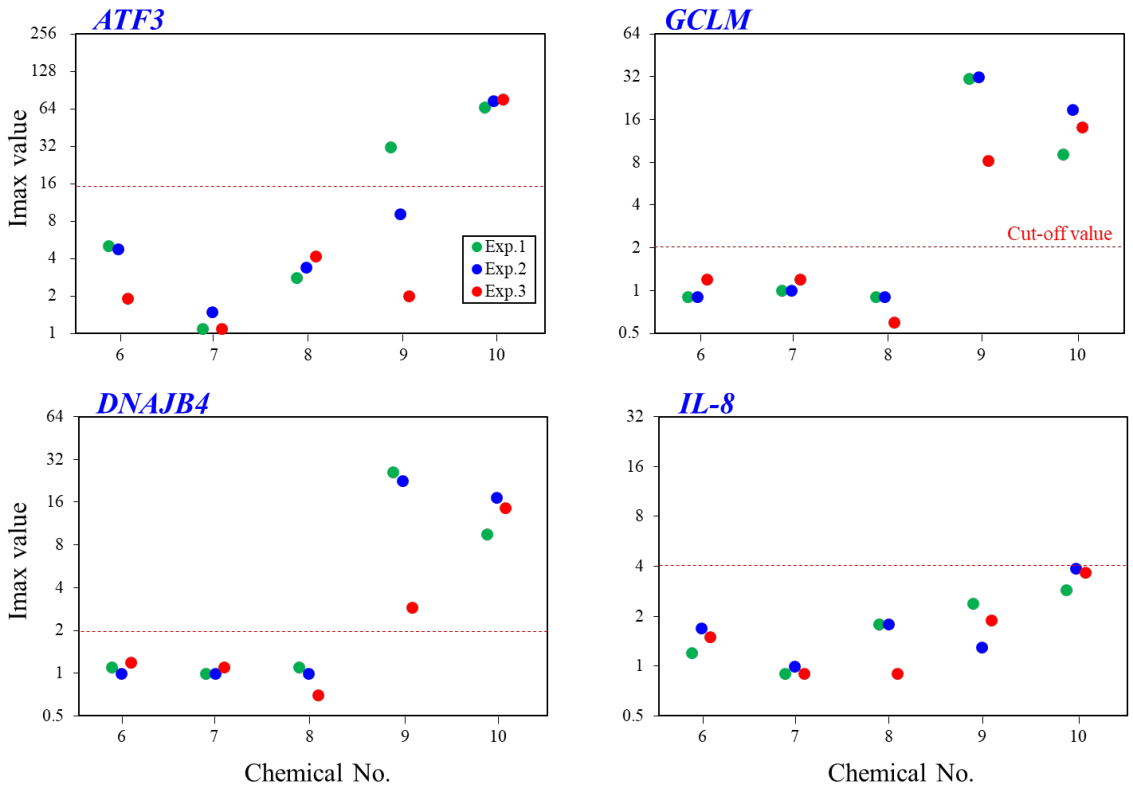


Figure 16. Distribution of the mean Imax values (n=3) for chemicals #6-10, as assessed by three individual experiments at FDSC. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

2-4. Revision of the SOP and other consideration

Several slight revisions (e.g., solubilization procedure, LDH assay protocol, note on topical application, and cDNA synthesis method) were proposed and approved at a web

meeting of the VMT (that took place on the 12th of May 2020) after the end of the Phase I-B study. These modifications were thus included in the SOP version 2.3.

3. Phase I-C

3-1. LION

a. Reproducibility (concordance in prediction)

Four sensitizers and one non-sensitizer were tested three times each. The same prediction in all three independent experiments was obtained for 5 out of the 5 chemicals tested, resulting in a WLR of 100% (Table 14).

Table 14. The results of Phase I-C, as conducted by LION.

No.	Chemical	Exp.1		Exp.2		Exp.3		Agreement between experiment
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
11	Lactic acid	DW	N	DW	N	DW	N	Yes
12	p-Phenylenediamine	AOO	P	AOO	P	AOO	P	Yes
13	Methyl heptine carbonate	AOO	P	AOO	P	AOO	P	Yes
14	Abietic acid	AOO	P	AOO	P	AOO	P	Yes
15	Farnesol	AOO	P	AOO	P	AOO	P	Yes

b. Reproducibility of the Imax value

Figure 17 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. The concordant results were demonstrated at 4/5, 5/5, 5/5, and 5/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, the Imax values for each of the marker genes had good consistency; a total of 19/20 demonstrated concordance across the three independent experiments performed.

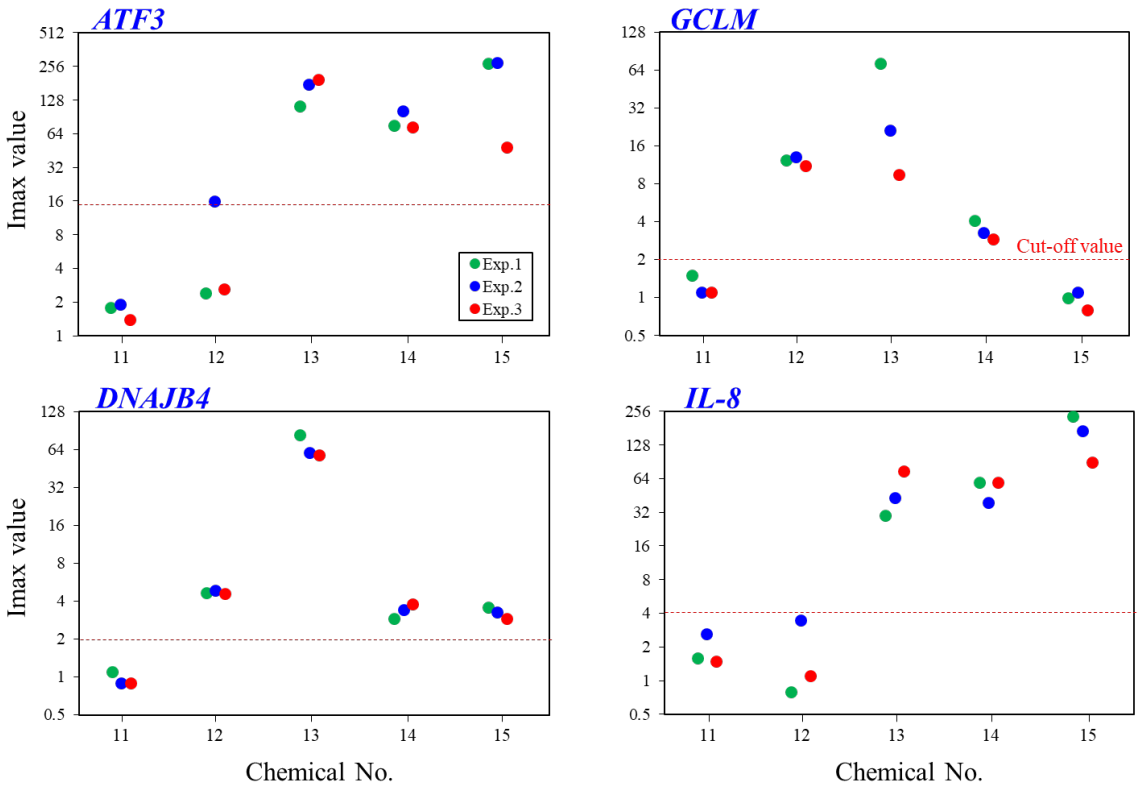


Figure 17. Distribution of the mean Imax values (n=3) for chemicals #11-15, as assessed by three individual experiments at LION. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

3-2. KOSÉ

a. Reproducibility (concordance in prediction)

The same prediction was obtained by the three independent experiments for 4 out of the 5 chemicals tested, resulting in a WLR of 80%. The chemical for which the classification was not concordant across all experiments was lactic acid (chemical #11) (Table 15).

Table 15. The results of Phase I-C, as conducted by KOSÉ.

No.	Chemical	Exp.1		Exp.2		Exp.3		Agreement between experiment
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
11	Lactic acid	DW	N	DW	N	DW	P	No
12	p-Phenylenediamine	AOO	P	AOO	P	AOO	P	Yes
13	Methyl heptine carbonate	AOO	P	AOO	P	AOO	P	Yes
14	Abietic acid	AOO	P	AOO	P	AOO	P	Yes
15	Farnesol	AOO	P	AOO	P	AOO	P	Yes

b. Reproducibility of the Imax value

Figure 18 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. In this case, it was considered that several chemicals (namely, the lactic acid at the 3rd experiment and farnesol at the 1st experiment) were affected by a cross-contamination with a volatile chemical, methyl heptine carbonate (MHC; chemical #13) and, as a result, it was assumed that their Imax values on marker genes (mainly on *GCLM* and *DNAJB4*) might have been over- or underestimated (see section 3-4). However, the Imax values for each of the marker genes had good consistency across all experiments in terms of whether they exceed the cut-off value or not; the concordant results were demonstrated at 4/5, 3/5, 4/5, and 4/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, a total of 15/20 demonstrated concordance across the three independent experiments performed.

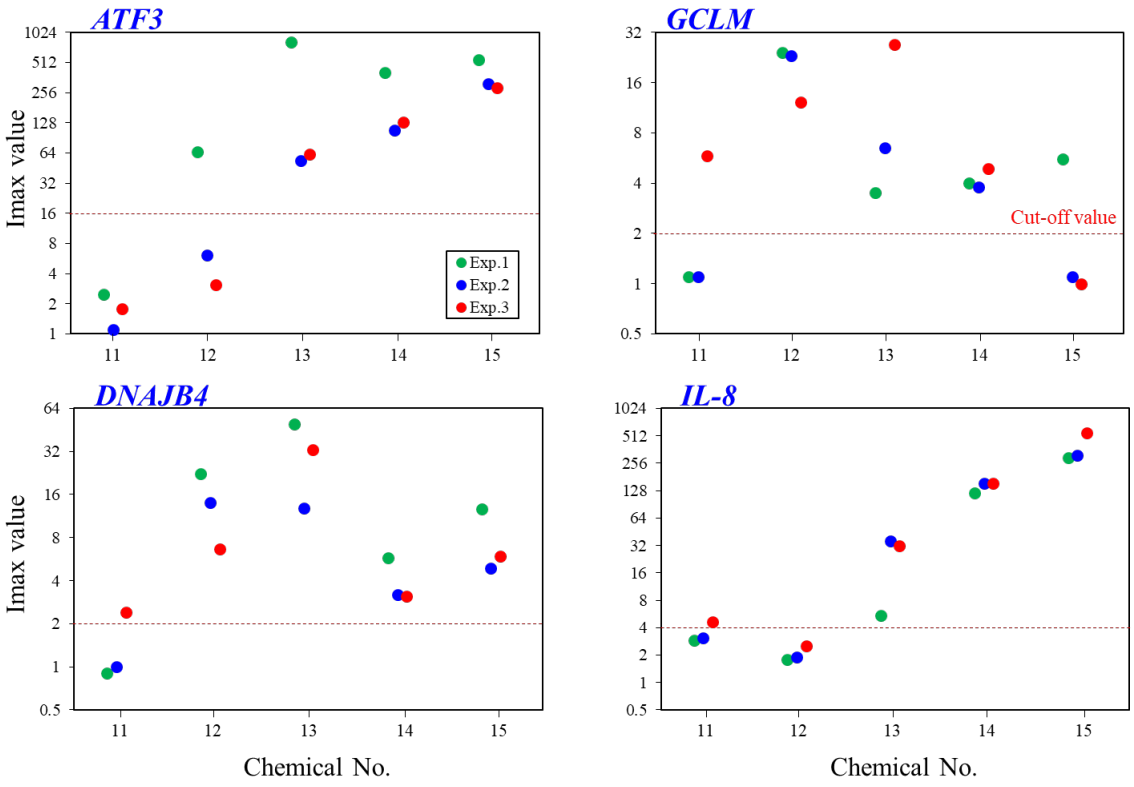


Figure 18. Distribution of the mean Imax values (n=3) for chemicals #11-15, as assessed by three individual experiments at KOSÉ. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

3-3. FDSC

a. Reproducibility (concordance in prediction)

The same prediction was obtained by the three independent experiments for 4 out of the 5 chemicals tested, resulting in a WLR of 80%. The chemical for which the classification was not concordant across all experiments was lactic acid (chemical #11) (Table 16).

Table 16. The results of Phase I-C, as conducted by FDSC.

No.	Chemical	Exp.1		Exp.2		Exp.3		Agreement between experiment
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
11	Lactic acid	DW	N	DW	N	DW	P	No
12	p-Phenylenediamine	50% EtOH	P	50% EtOH	P	50% EtOH	P	Yes
13	Methyl heptine carbonate	AOO	P	AOO	P	AOO	P	Yes
14	Abietic acid	AOO	P	AOO	P	AOO	P	Yes
15	Farnesol	AOO	P	AOO	P	AOO	P	Yes

b. Reproducibility of the Imax value

Figure 19 presents the distribution of the Imax values as assessed through three independent experiments for each of the 5 chemicals tested. In this case, it was considered that several chemicals (namely, the lactic acid at the 3rd experiment, p-phenylenediamine at the 2nd experiment, abietic acid at the 1st and 2nd experiment, as well as farnesol at the 1st experiment) were affected by a cross-contamination with a volatile chemical, MHC and, as a result, it was assumed that their Imax values on marker genes (mainly on *GCLM* and *DNAJB4*) might have been over- or underestimated (see section 3-4). However, for each of the marker genes, the Imax values had good consistency across all experiments in terms of whether they exceed the cut-off value or not; the concordant results were demonstrated at 4/5, 2/5, 3/5, and 5/5 chemicals in *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*, respectively. Altogether, a total of 14/20 demonstrated concordance across the three independent experiments performed.

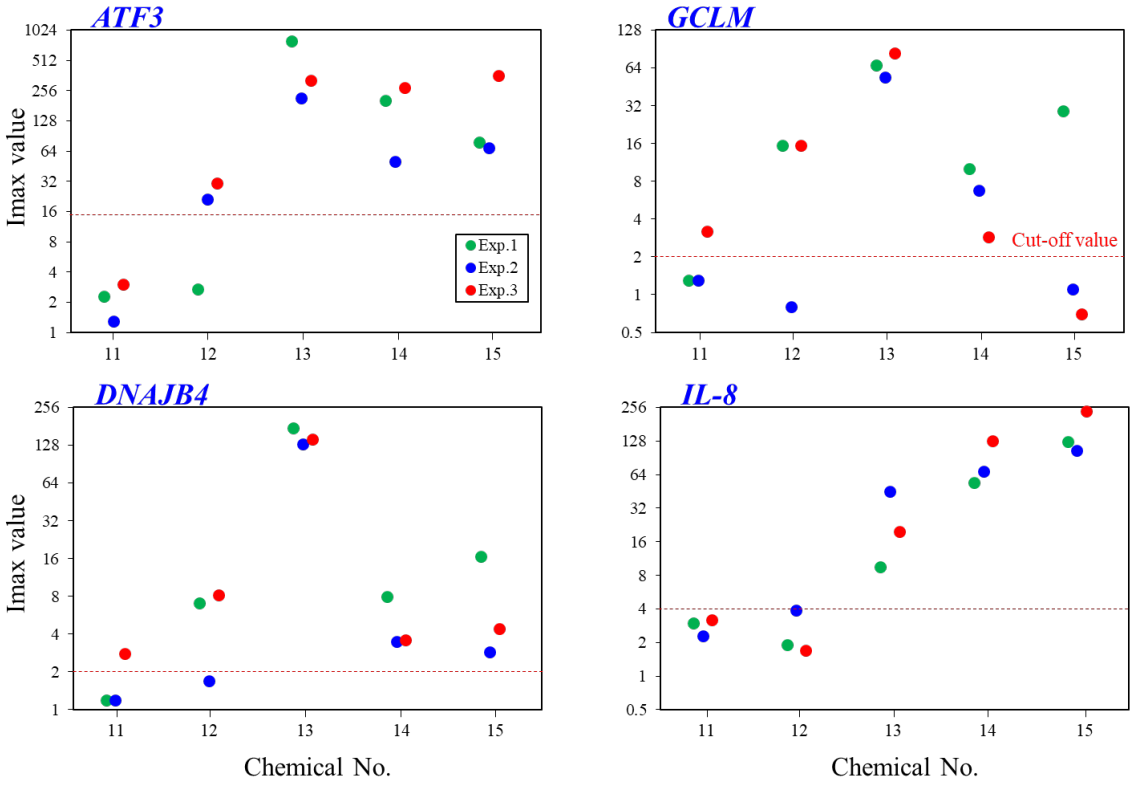


Figure 19. Distribution of the mean Imax values (n=3) for chemicals #11-15, as assessed by three individual experiments at FDSC. Dashed lines indicate the respective cut-off values for the induced expression of each marker gene.

3-4. Revision of the SOP and other consideration

The following revision was proposed, discussed and approved at a web meeting of the VMT (that took place on the 23rd of September 2020) after the end of the Phase I-C study; this modification was, thus, included in the SOP version 2.4.

- A cautionary note regarding the liquid chemical exposure was added. In order to avoid cross-contamination by volatile compounds, the tissue units that are used for liquid test chemicals should be kept separated from other test chemicals and controls (e.g., positive controls and vehicle controls) into individual 24-well plates. The detail is provided in Appendix 11.

4. Overall result of WLR and other considerations

4-1. WLR

Table 17 summarizes the overall results of Phase I for all of the 15 test chemicals that were assessed as part of this phase. When the results of retested experiments were adopted for benzisothiazolinone and sodium lauryl sulfate, the WLR was 93.3% (14/15) at LION and KOSÉ, and 86.7% (13/15) at FDSC.

Table 17. Overall results of Phase I for all of the 15 test chemicals assessed (N: negative; P: positive; / indicates original and retest).

No.	Chemical	LION			KOSÉ			FDSC		
		Exp.1	Exp.2	Exp.3	Exp.1	Exp.2	Exp.3	Exp.1	Exp.2	Exp.3
1	Glyoxal 40% solution in water	P	P	P	P	P	P	P	P	P
2	Lauryl gallate	N	N	N	N	N	N	N	P	N
3	Benzisothiazolinone	N/P	N/N	N/N	P	P	N/P	P	P	N/P
4	Diethylphthalate	P	P	P	P	P	P	P	P	P
5	Sodium lauryl sulfate	P	P	P	P	N/P	P	P	P	P
6	Hexane	N	N	N	N	N	N	N	N	N
7	Dextran	N	N	N	N	N	N	N	N	N
8	Tween 80	N	N	N	N	N	N	N	N	N
9	Ethyl acrylate	P	P	P	P	P	P	P	P	P
10	2,4-Dinitrochlorobenzene	P	P	P	P	P	P	P	P	P
11	Lactic acid	N	N	N	N	N	P	N	N	P
12	p-Phenylenediamine	P	P	P	P	P	P	P	P	P
13	Methyl heptine carbonate	P	P	P	P	P	P	P	P	P
14	Abietic acid	P	P	P	P	P	P	P	P	P
15	Farnesol	P	P	P	P	P	P	P	P	P

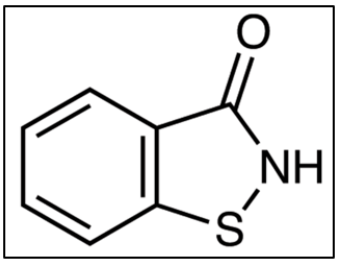
4-2. Discussion of discordant results

Benzisothiazolinone

Benzisothiazolinone was predicted by KOSÉ and FDSC to be a sensitizer, but it was predicted to be a non-sensitizer by LION. As shown in Figure 20, the Imax values for benzisothiazolinone were close to the cut-off values, and as a result it was predicted to be a sensitizer (on one occasion) by LION. Therefore, this borderline response might make it difficult to obtain concordant results in a consistent manner.

Table 18. Chemical structure of benzisothiazolinone, and its prediction at the three participating laboratories.

Benzisothiazolinone	LION	KOSÉ	FDSC
1st	P	P	P
2nd	N	P	P
3rd	N	P	P
prediction	N	P	P
solvent	AOO	AOO	AOO



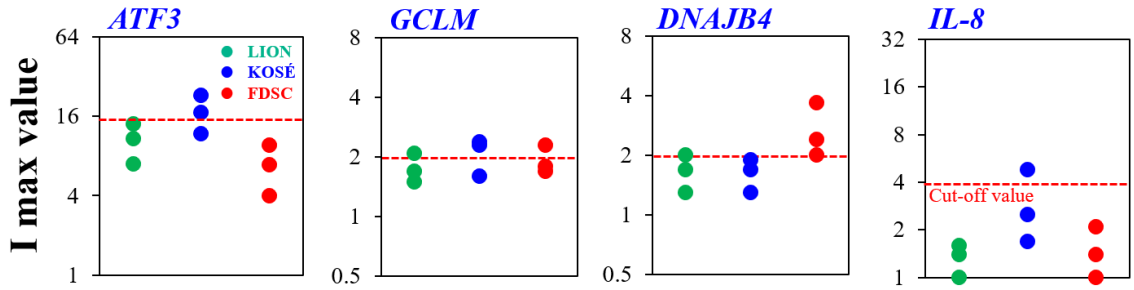


Figure 20. Distribution of the mean Imax values (n=3) for benzisothiazolinone.

4-3. Consideration around the positive controls

a. LION

Table 19 summarizes the results of the positive controls that were obtained during Phase I. As per requirement for a qualified testing, the mean cell viability of both positive controls should be equal to or greater than 80%. Furthermore, in the case of 0.78% w/v clotrimazole, the mean values of the fold induction of the *ATF3* and *IL-8* should exceed the respective cut-off values (*ATF3*: by 15-fold; *IL-8*: by 4-fold). Moreover, in the case of 0.10% w/v 4NBB, the mean values of the fold induction of the *GCLM* and *DNAJB4* should exceed the respective cut-off values (*GCLM*: by 2-fold; *DNAJB4*: by 2-fold). If these requirements are not met, the experiment should be discarded. In Phase I-A, the fold inductions of *ATF3* as a result of an exposure to 0.78% w/v clotrimazole did not exceed the cut-off value on two occasions. The precise cause of this remains unclear, but we suspect that the operator might have not been completely proficient at the early phases of the validation study. Regarding Phase I-B, the fold induction of the *ATF3* as a result of an exposure to 0.78% w/v clotrimazole did not exceed the cut-off value. For this case,

LION suggested that clotrimazole might have not been completely dissolved, and that the concentration of clotrimazole tested might have been lower than 0.78% w/v.

Table 19. Summary of the results obtained for the positive controls, during Phase I, by LION. If a gene expression fold induction did not meet the respective requirement criterion, then the value was highlighted in blue color.

		0.78 w/v% Clotrimazole					0.10 w/v% 4-Nitrobenzyl bromide				
	No.	*ATF3* (>15-fold)	*GCLM*	*DNAJB4*	*IL-8* (>4-fold)	Viability (≥80%)	*ATF3*	*GCLM* (>2-fold)	*DNAJB4* (>2-fold)	*IL-8*	Viability (≥80%)
Phase I-A	1	68.4	1.0	2.4	18.4	100.5	91.7	7.8	8.7	4.5	96.5
	2	17.1	0.8	1.3	6.7	103.2	69.7	4.9	6.1	4.0	98.0
	3	61.6	1.2	3.2	39.0	101.8	45.5	7.8	8.2	4.1	90.6
	4	7.4	0.6	1.4	7.1	105.5	17.8	3.2	3.8	1.3	102.7
	5	80.5	1.1	4.1	40.9	99.6	48.9	5.5	6.9	3.5	97.1
	6	33.5	0.8	1.8	9.5	101.5	56.0	4.3	4.9	1.5	100.3
	7	36.2	1.3	1.8	16.4	102.3	44.8	5.9	6.3	2.6	95.8
	8	39.7	1.1	6.1	24.3	98.2	14.0	4.4	5.4	1.3	97.4
	9	13.9	1.0	1.4	4.4	101.5	16.9	6.0	6.9	1.6	95.6
	10	19.0	0.9	2.7	13.6	109.3	4.5	4.8	4.5	1.8	106.0
	11	41.1	1.1	2.4	13.0	98.8	11.2	4.7	5.3	1.7	94.7
	12	26.2	1.1	1.6	12.8	97.7	57.0	8.5	13.8	3.3	97.4
	13	40.0	1.2	2.5	19.7	99.9	42.0	8.3	11.1	2.5	97.8
	14	55.3	1.1	3.0	16.1	98.5	48.8	6.9	9.5	2.8	96.4
Phase I-B	15	28.4	1.0	2.3	12.5	99.2	38.5	8.9	11.7	6.0	90.3
	16	18.5	0.9	1.8	14.5	98.0	41.0	6.2	12.2	5.9	90.1
	17	35.2	1.1	1.6	18.9	95.5	86.5	6.4	10.8	4.4	96.3
	18	6.8	0.9	1.3	4.0	98.4	30.7	7.5	9.1	3.1	101.8
	19	18.2	0.7	1.4	8.5	99.5	73.9	13.6	17.5	8.6	98.2
	20	35.1	1.1	1.5	14.5	99.7	15.5	3.6	4.7	3.2	97.2
Phase I-C	21	35.0	1.2	1.7	28.4	100.0	24.0	6.4	7.5	6.0	98.4
	22	64.5	1.1	3.6	46.3	94.7	31.1	9.2	11.7	6.8	94.6
	23	36.4	1.1	3.0	23.8	98.7	15.6	1.6	3.1	2.9	93.2
	24	26.9	0.4	1.5	16.3	95.7	16.3	0.6	1.8	3.1	91.4
	25	67.1	0.9	3.7	42.9	94.0	25.3	4.1	9.2	5.0	96.5
	26	49.9	1.0	2.1	57.0	99.5	19.0	3.3	5.2	5.9	97.1
	27	29.9	1.0	2.9	15.3	89.3	15.0	3.7	5.9	2.2	97.2
	28	22.2	0.9	1.4	36.2	99.1	15.6	4.4	7.4	4.5	95.4
	29	26.3	0.9	1.5	18.4	98.9	15.0	6.8	8.1	5.3	96.2
	30	31.8	1.3	3.9	30.8	100.8	7.5	3.3	6.9	2.7	92.0
	31	30.4	1.0	2.4	19.6	95.8	29.1	9.4	14.0	10.8	96.9

b. KOSÉ

Table 20 summarizes the results of the positive controls obtained during Phase I, at KOSÉ. In all experiments, the requirements of positive controls were fulfilled.

Table 20. Summary of the results obtained for the positive controls, during Phase I, by KOSÉ. If a gene expression fold induction did not meet the respective requirement criterion, then the value was highlighted in blue color.

	No.	0.78 w/v% Clotrimazole					0.10 w/v% 4-Nitrobenzyl bromide				
		ATF3 (>15-fold)	*GCLM*	*DNAJB4*	*IL-8* (>4-fold)	Viability (≥80%)	*ATF3*	*GCLM* (>2-fold)	*DNAJB4* (>2-fold)	*IL-8*	Viability (≥80%)
Phase I-A	1	97.1	0.9	3.0	39.1	100.4	22.0	3.3	4.8	1.6	97.7
	2	68.2	0.9	2.2	25.6	100.4	33.7	4.3	6.3	1.9	97.8
	3	59.0	1.0	2.4	21.0	99.6	33.4	5.8	8.7	2.7	97.7
	4	33.5	0.8	1.5	19.4	104.0	27.3	5.2	5.8	3.8	99.3
	5	60.4	1.0	2.8	30.9	99.4	25.4	5.2	8.1	2.9	96.9
	6	56.0	1.1	2.0	17.9	100.5	37.9	5.1	7.5	3.1	98.4
	7	72.2	0.8	2.1	26.0	100.4	24.1	3.9	6.3	2.4	97.4
	8	46.8	1.0	1.6	14.6	99.2	32.2	4.3	5.3	2.2	97.4
	9	79.0	0.9	1.8	42.6	101.2	38.2	3.8	4.6	3.7	98.2
Phase I-B	10	149.8	1.0	4.1	59.5	95.2	82.7	4.3	5.7	8.5	95.6
	11	53.7	1.1	2.2	34.3	100.7	32.6	2.7	4.0	4.0	93.9
	12	31.7	1.0	1.4	11.7	101.4	39.7	8.1	10.0	4.4	97.5
	13	41.5	1.1	2.1	21.0	101.1	35.8	5.1	7.3	4.4	97.1
	14	61.3	1.0	3.6	29.2	99.1	45.3	5.5	7.4	2.8	94.6
	15	26.0	1.1	1.6	23.3	100.3	22.9	4.0	5.5	3.4	92.1
Phase I-C	16	71.9	1.1	3.3	42.9	100.1	35.1	5.5	8.4	5.5	94.5
	17	28.8	0.8	1.7	27.8	101.1	24.8	4.7	6.1	4.3	98.3
	18	60.4	1.1	3.5	34.2	100.0	106.7	14.3	17.9	20.2	98.1
	19	19.2	0.8	1.7	44.0	99.0	20.5	5.2	7.8	7.3	96.0
	20	45.4	0.8	3.4	66.7	96.3	22.0	5.3	9.1	3.8	92.9
	21	112.1	1.4	4.5	109.9	97.5	21.0	2.1	4.1	5.4	91.8
	22	24.4	0.8	1.9	26.6	99.5	23.1	5.4	8.5	5.4	95.3
	23	22.4	0.9	1.5	26.4	101.9	14.6	3.8	5.2	5.6	97.6

c. FDSC

Table 21 summarizes the results of the positive controls that were obtained during Phase I, at FDSC. Phase I-A has shown that the fold induction on *ATF3* expression as a result of an exposure to 0.78% w/v clotrimazole did not exceed the cut-off value. The precise cause of that remains unclear, but a possible explanation might be that the operator was not completely proficient in the early phase of the validation study.

Table 21. Summary of the results obtained for the positive controls, during Phase I, by FDSC. If a gene expression fold induction did not meet the respective requirement criterion, then the value was highlighted in blue color.

		0.78 w/v% Clotrimazole					0.10 w/v% 4-Nitrobenzyl bromide				
	No.	*ATF3* (>15-fold)	*GCLM*	*DNAJB4*	*IL-8* (>4-fold)	Viability (≥80%)	*ATF3*	*GCLM* (>2-fold)	*DNAJB4* (>2-fold)	*IL-8*	Viability (≥80%)
Phase I-A	1	73.7	0.8	4.7	22.2	102.9	10.5	4.3	6.9	1.4	99.4
	2	4.7	0.6	1.1	4.1	98.4	10.2	2.9	4.4	3.0	99.9
	3	75.2	0.7	2.7	29.2	100.3	50.2	5.3	9.9	2.5	96.6
	4	84.8	0.9	4.3	34.9	100.0	19.1	5.3	6.2	2.7	97.2
	5	85.5	0.8	3.2	25.5	100.2	24.9	3.0	5.3	2.1	98.9
	6	128.6	1.0	4.5	29.7	100.2	41.4	4.8	6.4	2.3	98.5
	7	38.6	2.8	0.9	13.9	98.8	51.2	14.4	8.4	2.0	95.4
Phase I-B	8	252.6	1.0	11.0	69.5	91.0	100.3	8.4	18.1	6.7	93.3
	9	147.4	1.0	4.8	52.4	94.1	216.9	9.4	20.5	5.8	91.5
	10-1*	156.7	0.9	9.4	45.8	87.8	48.9	7.7	9.7	4.6	87.4
	10-2*	179.9	1.3	12.0	57.4	87.8	59.6	12.6	14.9	4.8	87.4
	11-1*	241.3	1.1	12.6	74.0	95.5	56.1	6.5	10.9	3.2	96.8
	11-2*	306.5	1.5	14.7	77.6	95.5	60.8	8.4	10.2	3.4	96.8
Phase I-C	12	149.0	0.9	14.2	63.2	89.4	48.6	16.6	22.4	5.2	87.2
	13-1*	123.1	1.2	11.7	117.7	97.6	49.9	8.1	15.0	4.0	97.8
	13-2*	53.9	1.3	7.1	55.3	97.6	25.3	7.5	11.7	2.1	97.8
	14	88.5	1.0	4.9	48.6	99.9	79.7	20.2	25.9	5.0	99.1

*: Positive controls were measured twice due to the fact that the gene expression measurement of the test chemicals that were assessed in main study No.10 (and also in No.11 and 13) were split into two parts.

5. Conclusion

The WLR for Phase I was 93.3% (14/15) at the experiments conducted by LION and KOSÉ, and 86.7% (13/15) at those conducted by FDSC; on average a WLR of 91.1% was achieved by the three participating laboratories. These results satisfied the 85% target value established by the VMT, thus indicating an excellent WLR for EpiSensA.

Between-Laboratory Reproducibility

As previously noted, several improvements to issues arising in Phase I were made to the SOP, and as a result, Phase II (Between-Laboratory Reproducibility) was performed using SOP version 2.4.

In order to minimize bias in the generation of the results for the assessment of the between-laboratory reproducibility, the participating laboratories were provided with a coded vial for each chemical. Participating laboratories had to perform a solubility check, vehicle determination and a dose-finding study prior to the performance of the main study for each chemical.

The between-laboratory reproducibility of the EpiSensA was assessed on the concordance of classification, positive (P) versus negative (N), among the three participating laboratories. Additionally, descriptive analyses were performed in order to evaluate the reproducibility of the three respective Imax values for the four marker genes.

1. Between-Laboratory Reproducibility (concordance in prediction)

The between-laboratory reproducibility was assessed based on test results obtained from 27 coded test chemicals. Final predictions for the 15 test chemicals that were subjected to three replicate tests during Phase I were based on the more frequent result.

1-1. Between-laboratory reproducibility of the 12 test chemicals used in Phase II

Since 15 test chemicals were subjected to three replicate tests during Phase I, and 12 test chemicals were subjected to single tests during Phase II, the results of Phase I were handled differently from those of Phase II. Thus, a preliminary evaluation of the between-laboratory reproducibility was made by using just the 12 test chemicals used for Phase II. Table 22 summarizes the predictions reached at each participating laboratory, and Figure 21 demonstrates the distribution of the Imax values (maximum fold-induction with more than or equal to 80% cell viability) for the 12 test chemicals used in Phase II. The between-laboratory reproducibility for these 12 test chemicals was calculated by using results from three laboratories, and was found to be 83.3%; a level that exceeded the minimum target value of 80% set by the VMT.

Table 22. Summary of the predictions reached at each participating laboratory for the 12 test chemicals used in Phase II.

No.	Chemical	LION		KOSÉ		FDSC		Agreement between laboratory
		Vehicle	Judge	Vehicle	Judge	Vehicle	Judge	
16	Acetanisole	AOO	P	AOO	N	AOO	P	No
17	1-Iodohexane	AOO	P	AOO	P	AOO	P	Yes
18	propylene glycol	DW	N	DW	P	DW	P	No
19	Benzyl butyl phthlate	AOO	N	AOO	N	AOO	N	Yes
20	Tetrachlorosalicylanilide	AOO	P	AOO	P	AOO	P	Yes
21	Isoeugenol	AOO	P	AOO	P	AOO	P	Yes
22	2-Aminophenol	AOO	P	AOO	P	50%EtOH	P	Yes
23	Glutaraldehyde	DW	P	DW	P	DW	P	Yes
24	Lilial	AOO	P	AOO	P	AOO	P	Yes
25	Methyl methacrylate	AOO	P	AOO	P	AOO	P	Yes
26	Amyl cinnamic aldehyde	AOO	P	AOO	P	AOO	P	Yes
27	Imidazolidinyl urea	DW	P	DW	P	DW	P	Yes

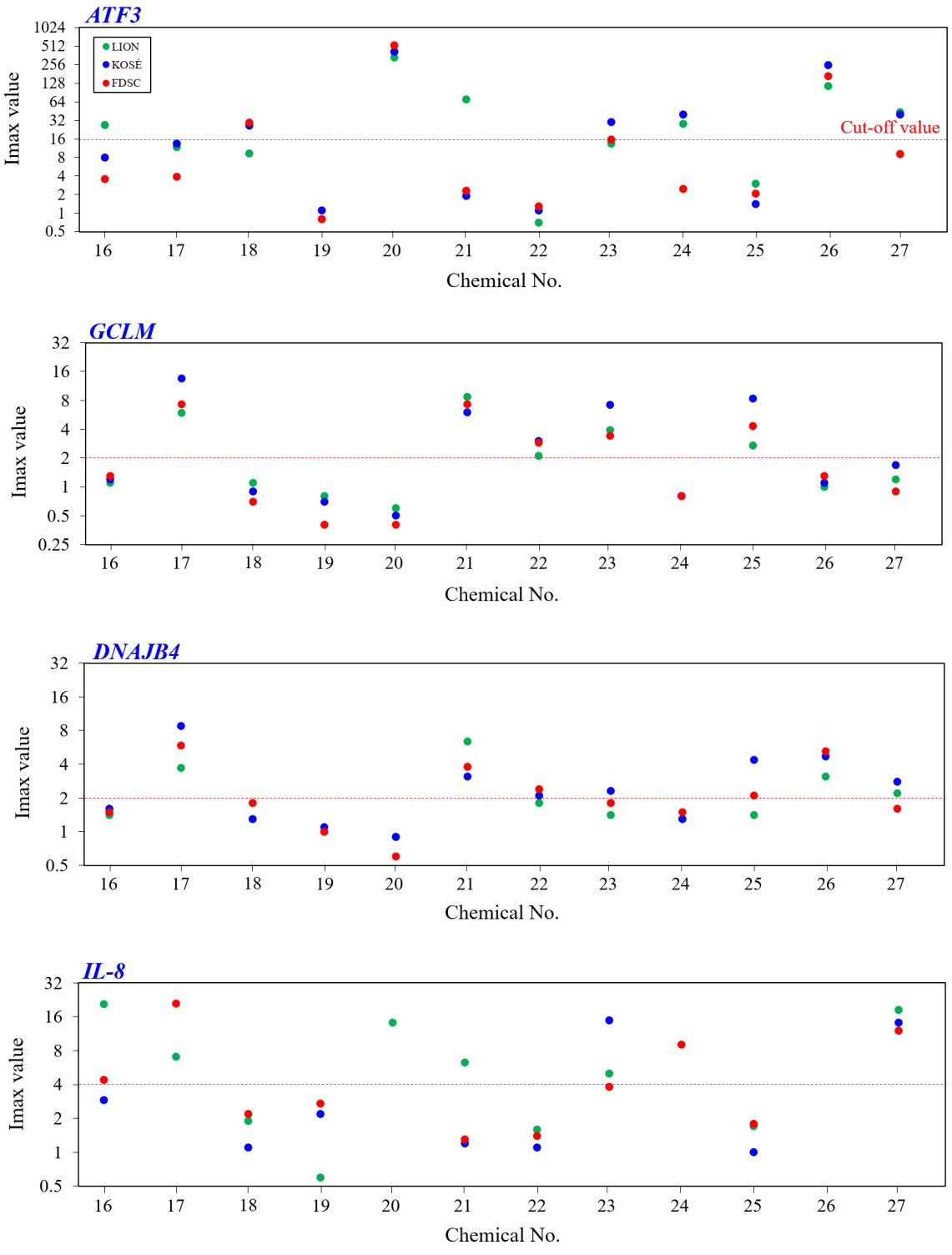


Figure 21. Distribution of the mean Imax values (n=3) for the 12 test chemicals used for Phase II. Dashed lines indicate the respective cut-off values for each marker gene.

1-2. Between-laboratory reproducibility of the 27 test chemicals used in Phases I and II

As shown in Table 23, the prediction of the sensitization potential for 24 of the 27 test chemicals was concordant across all three participating laboratories, and the between-laboratory reproducibility for these 27 test chemicals was 88.9%; a level that exceeded the minimum target value of 80% set by the VMT.

Table 23. Summary of the predictions reached at each participating laboratory regarding the 27 test chemicals used in Phases I and II.

No.	Chemical	LION	KOSÉ	FDSC	Agreement between laboratory
1	Glyoxal 40% solution in water	P	P	P	Yes
2	Lauryl gallate	N	N	N	Yes
3	Benzisothiazolinone	N	P	P	No
4	Diethyl phthalate	P	P	P	Yes
5	Sodium lauryl sulfate	P	P	P	Yes
6	Hexane	N	N	N	Yes
7	Dextran	N	N	N	Yes
8	Tween80	N	N	N	Yes
9	Ethyl acrylate	P	P	P	Yes
10	2,4-Dinitrochlorobenzene	P	P	P	Yes
11	Lactic acid	N	N	N	Yes
12	p-Phenylenediamine	P	P	P	Yes
13	Methyl heptine carbonate	P	P	P	Yes
14	Abietic acid	P	P	P	Yes
15	Farnesol	P	P	P	Yes
16	Acetanisole	P	N	P	No
17	1-Iodohexane	P	P	P	Yes
18	propylene glycol	N	P	P	No
19	Benzyl butyl phthlate	N	N	N	Yes
20	Tetrachlorosalicylanilide	P	P	P	Yes
21	Isoeugenol	P	P	P	Yes
22	2-Aminophenol	P	P	P	Yes
23	Glutaraldehyde	P	P	P	Yes
24	Lilial	P	P	P	Yes
25	Methyl methacrylate	P	P	P	Yes
26	Amyl cinnamic aldehyde	P	P	P	Yes
27	Imidazolidinyl urea	P	P	P	Yes

2. Other considerations

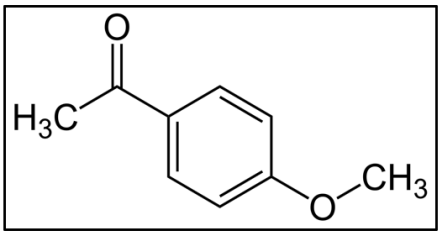
2-1. Discussion of non-concordant-BLR chemicals

Acetanisole

Acetanisole was predicted to be a non-sensitizer by KOSÉ, and a sensitizer by LION and FDSC (Table 24). As shown in Figure 22, the dose-response recorded by KOSÉ and LION were similar, but the acceptable test concentration (the concentration with more than or equal to 80% cell viability) used at LION was higher than that used at KOSÉ. For this reason, discordant results were confirmed between KOSÉ and LION. On the other hand, at FDSC, the fold induction of the *IL-8* expression exceeded the cut-off value at the lowest test concentration. The reason why that specific fold induction increased at the lower tested concentration is unclear.

Table 24. Chemical structure of acetanisole and overview of its prediction at the three participating laboratories.

Acetanisole	LION	KOSÉ	FDSC
prediction	P	N	P
solvent	AOO	AOO	AOO



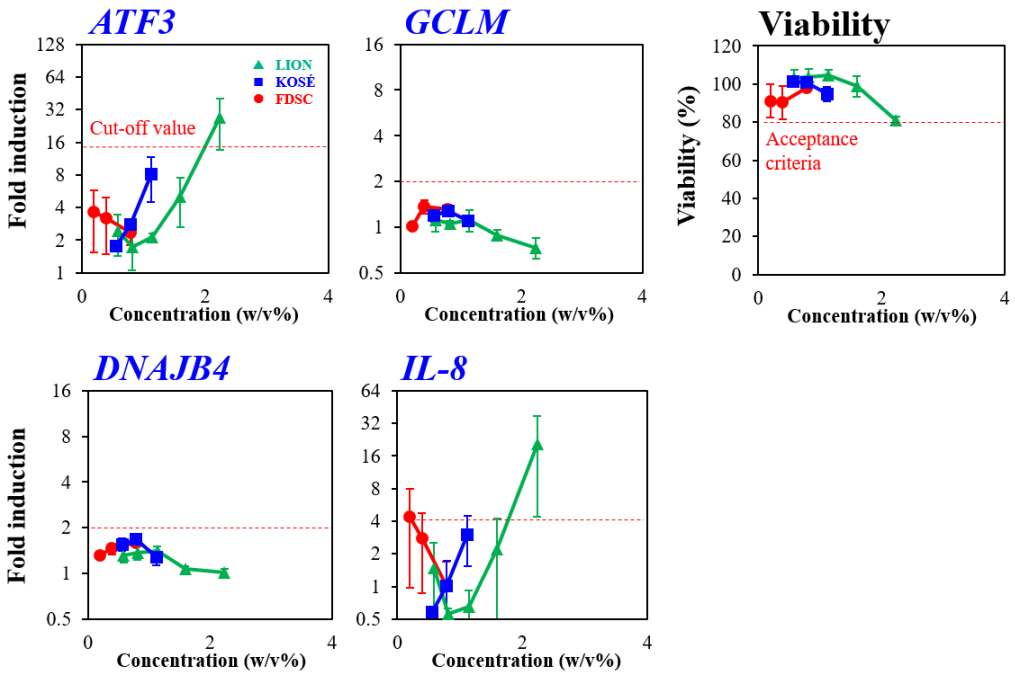


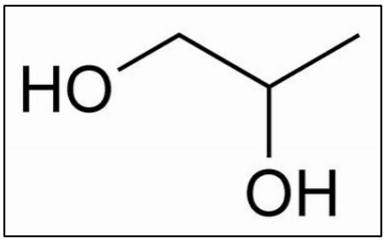
Figure 22. Dose-response of fold induction on four marker genes and viability when acetanisole was tested at the three participating laboratories (n=3, bars: S.D.).

Propylene glycol

Propylene glycol was predicted to be a non-sensitizer by LION, and a sensitizer by KOSÉ and FDSC (Table 25). As shown in Figure 23, the dose-response curves of the three participating labs were very similar, and the fold induction of the *ATF3* expression exceeded the cut-off value at 100% w/v concentration. However, the cell viability at 100% w/v concentration proved to be less than 80% at LION, and as such, the result at this concentration was discarded. This is the reason why discordant results were confirmed in the case of propylene glycol.

Table 25. Chemical structure of propylene glycol, and overview of its prediction at the three participating laboratories.

Propylene glycol	LION	KOSÉ	FDSC
prediction	N	P	P
solvent	DW	DW	DW



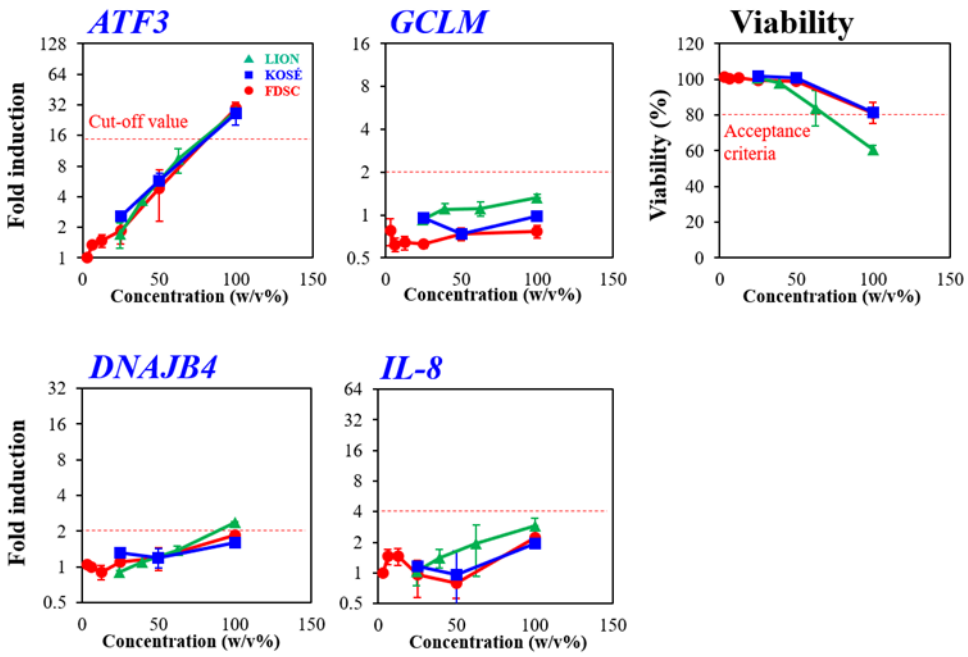


Figure 23. Dose-response of the fold induction for the four marker genes and of the cell viability when propylene glycol was tested at the three participating laboratories (n=3, bars: S.D.).

2-2. Consideration around the positive controls

a. LION

Table 26 summarizes the results of the positive controls that were obtained during Phase II by LION. The fold induction of the *ATF3* expression as a result of an exposure to 0.78% w/v clotrimazole did not meet the acceptance criteria set in three out of the nine experiments undertaken by LION. Therefore, the cause of failures was considered.

Table 26. Summary of positive controls that were obtained during Phase II by LION. When a fold induction did not meet the requirement criteria, its value was highlighted in blue.

No.	0.78 w/v% Clotrimazole					0.10 w/v% 4-Nitrobenzyl bromide				
	ATF3 (>15-fold)	*GCLM*	*DNAJB4*	*IL-8* (>4-fold)	Viability (≥80%)	*ATF3*	*GCLM* (>2-fold)	*DNAJB4* (>2-fold)	*IL-8*	Viability (≥80%)
1	29.2	1.0	1.4	23.8	104.7	16.1	3.7	6.8	4.1	103.5
2	10.3	2.9	4.1	9.0	99.3	5.8	4.8	7.2	2.6	99.1
3	28.9	1.0	2.5	20.5	103.7	11.0	5.6	8.5	3.1	98.0
4	4.4	0.7	1.6	5.0	100.6	5.2	4.8	6.7	1.8	92.2
5	32.5	0.9	1.0	35.9	100.6	19.8	5.4	10.3	5.2	97.1
6	29.3	1.0	2.0	23.9	99.9	14.8	3.6	4.8	1.6	96.0
7	47.3	0.9	2.3	43.9	99.5	16.9	4.8	6.1	4.7	95.0
8	11.3	0.9	3.3	14.5	96.4	1.8	3.1	4.4	0.7	91.7
9	22.5	0.8	2.4	12.4	97.5	12.6	3.9	7.4	1.4	96.8

Table 27 summarizes the *ATF3* expression Ct values for the AOO vehicle control, 0.78% w/v clotrimazole and a representative of test chemical (1-iodohexane) that was tested at both accepted experiments (No.5) and failed in one (No.4). At the failed experiment, the *ATF3* Ct values for the AOO control (28.41 to 30.65) were lower than those of the accepted experiment (30.81 to 31.46), and this means that the *ATF3* expression as a result of the exposure to the AOO control at the failed experiment was higher than that of the accepted one. Moreover, the *ATF3* expression Ct values for the AOO control (28.41 to 30.65) were also lower than those of the test chemical at the same experiment (30.20 to 31.82). From these results, one concludes that the *ATF3* expression would increase as a result of an exposure to the AOO control of the failed experiment, and as a result, the *ATF3* expression induction for 0.78% w/v clotrimazole would be underestimated.

Table 27. *ATF3* expression Ct values for the AOO vehicle control, for 0.78% w/v clotrimazole, as well as for the representative of test chemical (1-iodohexane).

Chemical	Conc. (w/v%)	Ct value of *ATF3*	
		Exp. No.5 (accepted)	Exp. No.4 (failed)
AOO (vehicle control)	-	30.81	**28.41**
		31.35	**29.29**
		31.46	**30.65**
Clotrimazole (positive control)	0.78	26.17	27.35
		26.41	26.40
		26.10	26.13
1-Iodohexane	13.3	Not tested	31.39
			31.82
			30.20
	18.6	30.72	31.54
		30.91	31.82
		30.06	31.69
	26.0	31.41	31.30
		29.55	31.07
		29.82	31.23

LION provided some comments on their failed experiments. Figure 24 presents an example plate layout for the 4th experiment conducted by LION. At the failed experiments, LION said, a high concentration of the test chemical might be contained at the position of the killed control which is usually set next to the AOO control. They said that at the three failed experiments, the test chemical applications at a high concentration on several RhEs failed and had to be performed on fresh RhEs again. Moreover, failed RhEs with a high concentration of the test chemical might have been re-used as killed controls by adding 10% Triton X-100 due to lack of fresh RhEs; this procedure is not allowed by the SOP and test chemicals should have not been contained in the killed control. Based on these comments, it was assumed that the volatile test chemicals contained in the killed control might have cross-contaminated the AOO control.

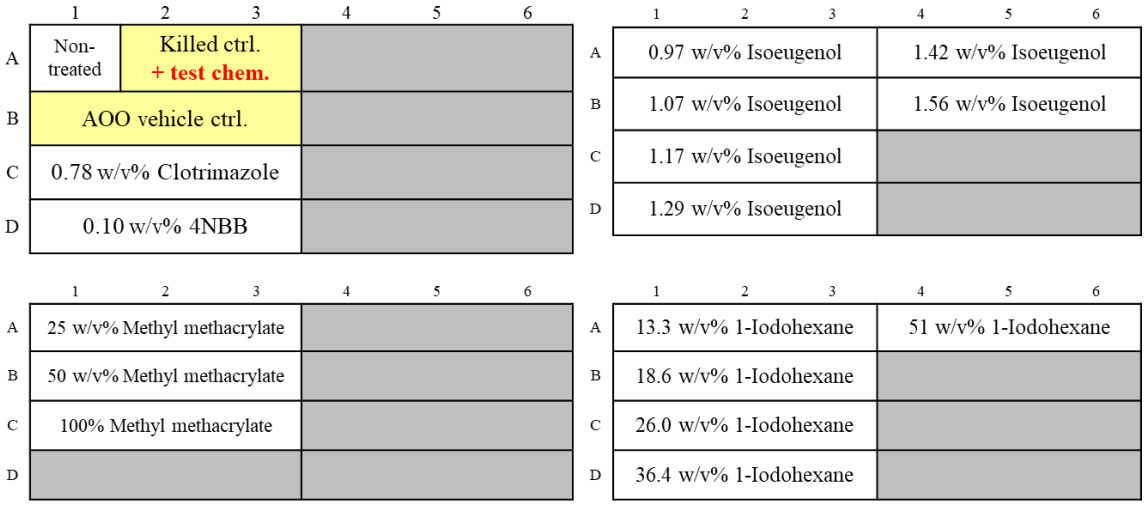


Figure 24. Plate layout of the 4th experiment conducted by LION.

Therefore, the cross-contamination potentials of the test chemicals were verified by the lead laboratory. In this case, three test chemicals (namely, methyl methacrylate, isoeugenol and 1-iodohexane) were applied at the failed experiments (isoeugenol and 1-iodohexane were applied at the 2nd experiment; methyl methacrylate, isoeugenol and 1-iodohexane were applied at the 4th experiment; 1-iodohexane was applied at the 8th experiment). In addition, all of them were in liquid form and were dissolved with AOO. Figure 25 demonstrates the plate layout used for the verification process and the possible results for an example chemical X and a chemical Y. At first, one of test chemicals was applied on RhEs, and 10% w/v triton X-100 was also applied on the same RhEs. After 6 h of incubation, the Ct value of the AOO control was measured. If chemical X was applied on the killed control and the *ATF3* expression Ct value of AOO was lower than the variation, then chemical X would demonstrate a cross-contamination effect. On the other hand, if chemical Y was applied on the killed control and the Ct value lied within or higher than the variation, than the chemical Y would not demonstrate any cross-contamination effect.

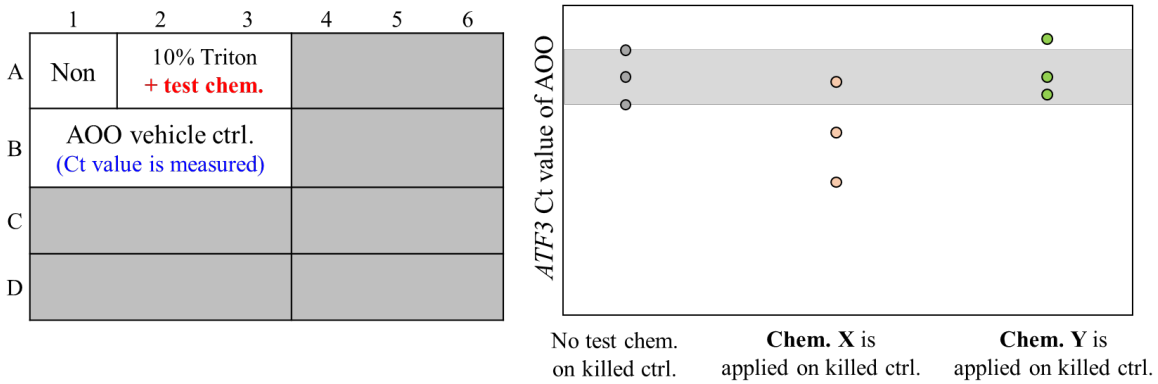


Figure 25. Plate layout used for the verification process, and possible results for chemicals X and Y. Gray area represents the variation of the AOO control when no test chemical was applied on the killed control.

Figure 26 presents the results of the verification performed. At the lead laboratory, almost all obtained Ct values were lower than the variation for the AOO control when no test chemical was applied on the killed control; as a result, all three chemicals would likely demonstrate a cross-contamination effect. In addition, the same experiment was performed by LION. The latter has suggested that methyl methacrylate and isoeugenol (and maybe isoeugenol; one of the three scored lower than the variation) have likely demonstrated a cross-contamination effect at these experiments. In summary, the cross-contamination of the test chemical was strongly suggested as the cause of failure observed at positive control.

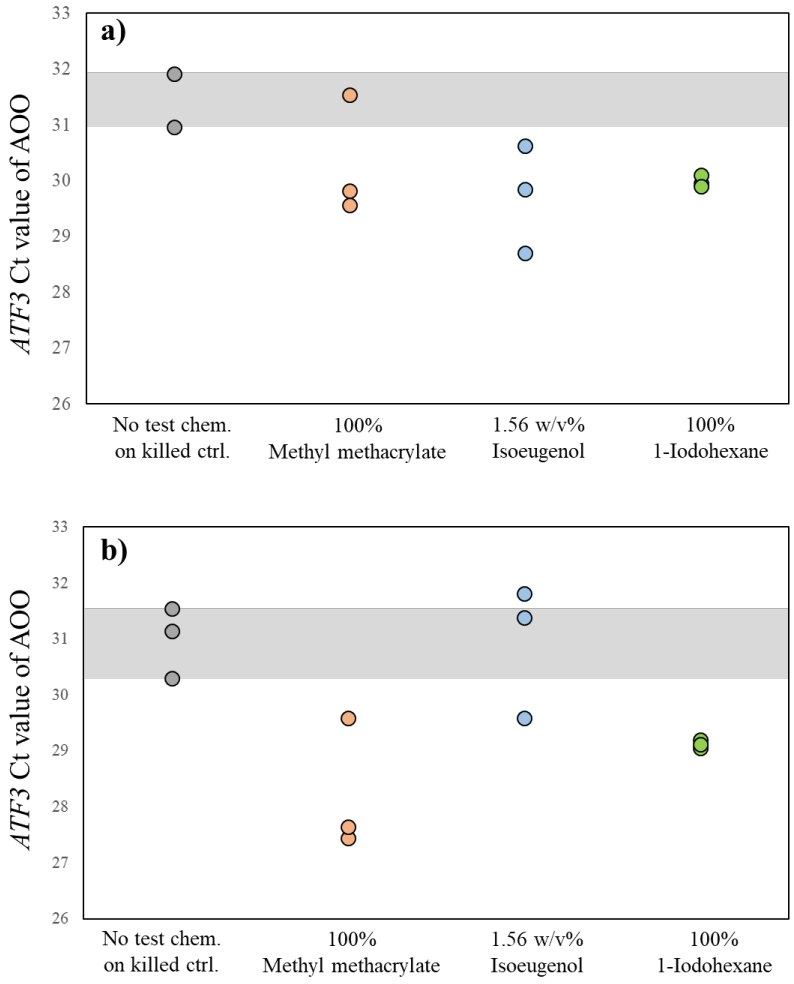


Figure 26. The verification of the cross-contamination potential of the test chemicals that were assessed at the failed experiments; (a): performed by lead laboratory; (b): performed by LION. Gray areas represent the variation for the AOO control when no test chemical was applied on the killed control.

b. KOSÉ

Table 28 summarizes the results of the positive controls that were obtained by KOSÉ during Phase II. The fold induction on *ATF3* expression by 0.78% w/v clotrimazole did not meet the acceptance criteria in one out of the six performed experiments. The VMT considered whether there was a significant difference between the fold induction values of the failed experiment at KOSÉ and those obtained by the lead laboratory. It was confirmed that a cross-contamination did not occur due to the fact that KOSÉ applied the killed control appropriately in its experiments. Firstly, the *ATF3* fold inductions of KOSÉ

lied within the variation of historical data (12.1 to 305.2-fold) that were obtained during the period of the EpiSensA validation study by the lead laboratory (21st of November 2018 ~ 12th of May 2021). Secondly, 0.78% w/v clotrimazole has been tested 31 times at KOSÉ during the validation study, and only failed to meet the criteria once. On the other hand, 0.78% w/v clotrimazole have been tested 95 times at the lead laboratory, and only failed to meet the criteria twice. Therefore, the frequencies of failure were comparable between KOSÉ and the lead laboratory. Regarding either the variation of the fold induction or the frequency of failure, there is no significant difference between KOSÉ and the lead laboratory. Additionally, the *ATF3* expression fold induction for KOSÉ at the failed experiment (14.9-fold) was very close to the cut-off value (15-fold).

Table 28. Summary of the positive controls that were obtained during Phase II at KOSÉ. If a fold induction did not meet the requirement criteria, its value was highlighted in blue color.

No.	0.78 w/v% Clotrimazole					0.10 w/v% 4-Nitrobenzyl bromide				
	ATF3 (>15-fold)	*GCLM*	*DNAJB4*	*IL-8* (>4-fold)	Viability (≥80%)	*ATF3*	*GCLM* (>2-fold)	*DNAJB4* (>2-fold)	*IL-8*	Viability (≥80%)
1	55.2	1.0	3.5	146.9	99.2	19.1	4.5	4.6	5.7	97.6
2	28.9	0.8	1.9	48.3	97.6	15.7	4.5	6.4	3.9	93.5
3	72.8	0.9	2.8	96.4	100.8	13.7	3.5	4.3	4.4	95.4
4	37.2	0.8	1.6	71.2	100.7	24.8	3.9	5.1	6.9	97.2
5	32.8	0.9	2.0	43.5	100.4	15.4	4.3	4.8	3.5	96.5
6	14.9	0.8	1.4	10.8	98.7	19.6	4.1	6.8	5.6	97.2
7	19.0	0.8	1.4	22.5	99.1	13.3	4.3	4.8	4.0	93.0
8	20.6	0.9	1.7	19.5	98.3	12.5	3.6	4.1	1.9	94.7

c. FDSC

Table 29 summarizes the results of the positive controls that were obtained during Phase II at FDSC. The positive controls met the acceptance criteria at all experiments.

Table 29. Summary of positive controls that were obtained during Phase II at FDSC. If a fold induction did not meet the requirement criteria, its value was highlighted in blue color.

No.	0.78 w/v% Clotrimazole					0.10 w/v% 4-Nitrobenzyl bromide				
	ATF3 (>15-fold)	*GCLM*	*DNAJB4*	*IL-8* (>4-fold)	Viability (≥80%)	*ATF3*	*GCLM* (>2-fold)	*DNAJB4* (>2-fold)	*IL-8*	Viability (≥80%)
1	170.4	1.0	8.3	238.5	99.0	222.4	10.4	26.8	22.5	95.0
2	97.5	0.8	3.4	71.4	98.7	43.7	4.0	8.1	3.5	97.0
3	54.2	0.7	2.2	85.4	98.2	55.7	7.6	13.8	13.7	87.1
4	147.6	0.7	5.8	168.9	98.7	98.0	8.1	22.8	15.9	96.7

d. Revision of the SOP

The following revision was proposed, discussed and approved at a web meeting of the VMT (that took place on the 13th of July 2021) after the end of the Phase II study; these modifications were, thus, included in the SOP version 2.5:

- The cautionary note regarding the cross-contamination of the liquid test chemicals (added to the SOP after Phase I-C) was transferred after the "topical application" section, and it was revised so as an operator can easily understand the effects of cross-contamination. In addition, the topical application procedure description was modified so as to prevent cross-contamination.

- The note on the topical application (including the relevant example cases) was revised.

3. Conclusion

The concordant predictions of the skin sensitization potential for the 27 test chemicals assessed by the three participating laboratories was used to measure the between-laboratory reproducibility. The between-laboratory reproducibility indicated a concordance for sensitizers and non-sensitizers of 88.9% when the final prediction for the chemicals that were tested 3 times in each laboratory was based on the median classification. The VMT considered the results as more than satisfactory since an 80% minimum target value for the between-laboratory reproducibility was specified in the EpiSensA study plan.

Quality Check

1. Chemical distribution

1-1. Chemical Acquisition, Coding and Distribution

The assessment of laboratory transferability, and within- and between-laboratory reproducibility and predictivity, in all test facilities were performed with the coded chemicals. The coding was supervised by JaCVAM (See Appendix 1).

1-2. Handling

The chemical master at each test facility received complete information considered essential regarding the test chemicals (physical state, weight or volume of sample, specific density for liquid test chemicals, and storage instructions) by JaCVAM. Moreover, the test facility chemical master stored each chemical at conditions in accordance with the storage instructions and received sealed safety information such as the Material Safety Data Sheet (MSDS) describing the hazards identification and exposure controls/personal protection for each chemical (See Appendix 2.1 and 2.2). The test chemicals were delivered directly to the study director and the study director was not shown the MSDSs. The study director was to refer to the MSDSs only in the event of an accident. If the study director referred to the MSDS, he/she was not to reveal the content of the MSDS to the test facility technicians.

No accidents occurred during the course of the validation study, and all test facilities returned the MSDSs for the test chemicals to JaCVAM in their sealed envelope upon completion of the validation study. All test chemicals were disposed of in compliance with the rules and regulations of the test facilities upon completion of the validation study.

2. Quality assurance

All the records (data sheets and record sheets) from the participating laboratories were checked by Dr. Hajime Kojima and Koji Yamakage, JaCVAM (See Appendix 3). The record sheets mean “Reagent records, solubility test, Cell culture records, Test records and data sheets”. They are total more than 300 pages and available at JaCVAM website (http://www.jacvam.jp/validation11-login.html). Testing performed as part of a validation study were carried out in non-GLP laboratories in accordance with the principles of GLP (OECD, 1998) and necessarily include, without being limited to, the use of protocol and adequate recording of data as well as suitable reporting of results and archival record keeping.

In the SOP, the incubation period is fixed for 6 hours. The QC team checked at the incubation periods based on all the record sheets and we found no acceptable one record which was obtained the data exposed for 6 hours and more 30 minutes at the Dose-finding study in Lab.C. However, the following note is addressed in the protocol (appendix 7, page 17). Therefore, we considered no data for the Dose-finding study is available and there is acceptable the results of the main study.

NOTE: Although it is possible to start the experiments for the main study by skipping the dose finding study, the results of the tested concentrations that show a mean cell viability of <80% should not be considered. The results of at least one tested concentration that shows ≥ 80% mean cell viability should be used to determine whether the test chemicals are positive or negative.

On the other hand, the QC team wondered whether there could be an impact of the length of the exposure period, which was not specified in the SOP. Therefore, we checked the exposure periods based on all the record sheets. In particular, we calculated from the start time of washing to last applied time of test chemicals excluding the dose finding study in Lab.C. As results, the following data were found.

Lab A: 5:40 to 6:04 hours

Lab B: 5:30 to 6:20 hours

Lab C: 5:48 to 6:26 hours

If the exposure periods are set up in the SOP, the QC team recommend the test developers to be added the acceptable exposure periods for 6 ± 0.5 hours.

In addition, we found several insufficient descriptions in the record sheets in all the laboratories. Especially we found inconsistency between the raw data and the records in the LDH assay and the RNA concentration measurement in all the laboratories, because a few record sheets were incomplete. At least, correct record sheets would be provided before the validation study and test developers and an independent organization should be checked them in advance.

Regarding a few data of the LDH assay and the RNA concentration measurement in the record sheets, we asked these points to the laboratories. As a result, the QC team decided no influence of misdescriptions for the validated outcome.

The QC team checked carefully the other results and judged all data within acceptable ranges. The QC team recommends JaCVAM that the formal validation study participated with GLP laboratories should be planned and done.

Predictive Capacity

1. Predictive capacity for LLNA data

Table 30 presents the skin sensitization potential of all the test chemicals as predicted by each of the three participating laboratories during the EpiSensA validation testing. Reference results are based on LLNA results which is available in Annex 2 of the Supporting document to the Guideline on Defined Approaches for Skin Sensitisation (OECD, 2021) except for dextran and tween 80.

With respect to the three replicate tests performed for the 15 test chemicals used in Phase I, the prediction of skin sensitization potential was made based on the two concordant findings.

Table 30. Predicted skin sensitization potential of all of the 27 test chemicals assessed at the three participating laboratories.

No.	Chemicals	Reference result	LION			KOSÉ			FDSC		
			Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3
1	DNCB	1A	P	P	P	P	P	P	P	P	P
2	Lauryl gallate	1A	N	N	N	N	N	N	N	P	N
3	p-Phenylenediamine	1A	P	P	P	P	P	P	P	P	P
4	Methyl heptine carbonate	1A	P	P	P	P	P	P	P	P	P
5	Glyoxal 40% solution in water	1A	P	P	P	P	P	P	P	P	P
6	Benzisothiazolinone	1B	P	N	N	P	P	P	P	P	P
7	Farnesol	1B	P	P	P	P	P	P	P	P	P
8	Ethyl acrylate	1B	P	P	P	P	P	P	P	P	P
9	Abietic acid	1B	P	P	P	P	P	P	P	P	P
10	Sodium lauryl sulfate	1B	P	P	P	P	P	P	P	P	P
11	Diethyl phthalate	No category	P	P	P	P	P	P	P	P	P
12	Hexane	No category	N	N	N	N	N	N	N	N	N
13	Dextran	No category[a]	N	N	N	N	N	N	N	N	N
14	Tween80	No category[a]	N	N	N	N	N	N	N	N	N
15	Lactic acid	No category	N	N	N	N	N	P	N	N	P
16	Tetrachlorosalicylanilide	1A	P			P			P		
17	Glutaraldehyde	1A	P			P			P		
18	2-Aminophenol	1A	P			P			P		
19	Isoeugenol	1A	P			P			P		
20	Lilial	1B	P			P			P		
21	Methyl methacrylate	1B	P			P			P		
22	Amyl cinnamic aldehyde	1B	P			P			P		
23	Imidazolidinyl urea	1B	P			P			P		
24	Acetanisole	No category	P			N			P		
25	1-Iodohexane	No category	P			P			P		
26	propylene glycol	No category	N			P			P		
27	Benzyl butyl phthalate	No category	N			N			N		

a: Urbisch et al., 2015

The predictive capacity for the results provided in Table 31 is summarized in Table 32. When compared to LLNA test results, the accuracy at each participating laboratory was

81.5% at LION, 85.2% at KOSÉ, and 81.5% at FDSC; the balanced accuracy was 77.8%, 80.6%, and 75.0%; the sensitivity was 88.9%, 94.4%, and 94.4%; and the specificity was 66.7%, 66.7%, and 55.6%, respectively.

Table 31. Predicted skin sensitization potential of the 27 test chemicals assessed at the three participating laboratories.

No.	Chemicals	Reference result	LION	KOSÉ	FDSC
1	DNCB	1A	P	P	P
2	Lauryl gallate	1A	N	N	N
3	p-Phenylenediamine	1A	P	P	P
4	Methyl heptine carbonate	1A	P	P	P
5	Glyoxal 40% solution in water	1A	P	P	P
6	Benzisothiazolinone	1B	N	P	P
7	Farnesol	1B	P	P	P
8	Ethyl acrylate	1B	P	P	P
9	Abietic acid	1B	P	P	P
10	Sodium lauryl sulfate	1B	P	P	P
11	Diethyl phthalate	No category	P	P	P
12	Hexane	No category	N	N	N
13	Dextran	No category	N	N	N
14	Tween80	No category	N	N	N
15	Lactic acid	No category	N	N	N
16	Tetrachlorosalicylanilide	1A	P	P	P
17	Glutaraldehyde	1A	P	P	P
18	2-Aminophenol	1A	P	P	P
19	Isoeugenol	1A	P	P	P
20	Lilial	1B	P	P	P
21	Methyl methacrylate	1B	P	P	P
22	Amyl cinnamic aldehyde	1B	P	P	P
23	Imidazolidinyl urea	1B	P	P	P
24	Acetanisole	No category	P	N	P
25	1-Iodohexane	No category	P	P	P
26	propylene glycol	No category	N	P	P
27	Benzyl butyl phthalate	No category	N	N	N

Table 32. Predictive capacity compared to LLNA for the 27 test chemicals at each of the three participating laboratories.

Cumulative

Reference result (LLNA)	+	−
+ (N=51)	**50**	**4**
− (N=30)	**10**	**17**

Total	
Sensitivity	92.6%
Specificity	63.0%
Acccuracy	82.7%
Balanced accuracy	77.8%

LION

	+	−
+	16	2
−	3	6

Total	
Sensitivity	88.9%
Specificity	66.7%
Acccuracy	81.5%
Balanced accuracy	77.8%

KOSÉ

	+	−
+	17	1
−	3	6

Total	
Sensitivity	94.4%
Specificity	66.7%
Acccuracy	85.2%
Balanced accuracy	80.6%

FDSC

	+	−
+	17	1
−	4	5

Total	
Sensitivity	94.4%
Specificity	55.6%
Acccuracy	81.5%
Balanced accuracy	75.0%

Thus, the cumulative accuracy for the 27 test chemicals used in Phases I and II at the three participating laboratories was 82.7%, the cumulative balanced accuracy was 77.8%, the cumulative sensitivity was 92.6%, and the cumulative specificity was 63.0%.

2. Predictive capacity for human data

Table 33 presents human data and predicted skin sensitization potential of the 27 test chemicals assessed at the three participating laboratories. The predictive capacity for the results provided in Table 33 is summarized in Table 34. Because human skin sensitization potential of acetanisole, 1-iodohexane, and benzyl butyl phthalate are unclear, these three chemicals are excluded from calculation of the predictive capacity. When compared to human data, the accuracy was 83.3% at the three participating laboratories; the sensitivity was 88.2% at LION, 94.1% at KOSÉ, and 94.1% at FDSC; and the specificity was 71.4%, 57.1%, and 57.1%, respectively.

Table 33. Human data and predicted skin sensitization potential of the 27 test chemicals assessed at the three participating laboratories.

No.	Chemicals	DASS human ref. data	Basketter human potency[a]	Reference result (human data)	LION	KOSÉ	FDSC
1	DNCB	1A	Cat.1	P	P	P	P
2	Lauryl gallate	NA	Cat.2	P	N	N	N
3	p-Phenylenediamine	1A	Cat.1	P	P	P	P
4	Methyl heptine carbonate	-	Cat.2	P	P	P	P
5	Glyoxal 40% solution in water	1A	Cat.2	P	P	P	P
6	Benzisothiazolinone	NA	Cat.2	P	N	P	P
7	Farnesol	1B	Cat.3	P	P	P	P
8	Ethyl acrylate	1B	Cat.4	P	P	P	P
9	Abietic acid	NA	Cat.3	P	P	P	P
10	Sodium lauryl sulfate	No category	Cat.6	N	P	P	P
11	Diethyl phthalate	NA	Cat.6	N	P	P	P
12	Hexane	No category	Cat.6	N	N	N	N
13	Dextran	-	Cat.6	N	N	N	N
14	Tween80	-	Cat.6	N	N	N	N
15	Lactic acid	NA	Cat.6	N	N	N	N
16	Tetrachlorosalicylanilide	1A	Cat.1	P	P	P	P
17	Glutaraldehyde	1A	Cat.2	P	P	P	P
18	2-Aminophenol	NA	Cat.2	P	P	P	P
19	Isoeugenol	1B	Cat.2	P	P	P	P
20	Lilial	1B	Cat.4	P	P	P	P
21	Methyl methacrylate	NA	Cat.4	P	P	P	P
22	Amyl cinnamic aldehyde	NA	Cat.4	P	P	P	P
23	Imidazolidinyl urea	1B	Cat.3	P	P	P	P
24	Acetanisole	NA	-	-	P	N	P
25	1-Iodohexane	NA	-	-	P	P	P
26	propylene glycol	No category	Cat.5	N	N	P	P
27	Benzyl butyl phthalate	NA	-	-	N	N	N

a: Basketter et al., 2014

NA: not available

Table 34. Predictive capacity compared to human data for the 24 test chemicals at each of the three participating laboratories.

Cumulative

Reference result (human data)	+	−
+ (N=51)	**47**	**4**
− (N=21)	**8**	**13**

Total	
Sensitivity	92.2%
Specificity	61.9%
Acccuracy	83.3%
Balanced accuracy	77.1%

LION

	+	−
+	15	2
−	2	5

Total	
Sensitivity	88.2%
Specificity	71.4%
Acccuracy	83.3%
Balanced accuracy	79.8%

KOSÉ

	+	−
+	16	1
−	3	4

Total	
Sensitivity	94.1%
Specificity	57.1%
Acccuracy	83.3%
Balanced accuracy	75.6%

FDSC

	+	−
+	16	1
−	3	4

Total	
Sensitivity	94.1%
Specificity	57.1%
Acccuracy	83.3%
Balanced accuracy	75.6%

Thus, the cumulative accuracy for the 24 test chemicals used in Phases I and II at the three participating laboratories was 83.3%, the cumulative balanced accuracy was 77.1%, the cumulative sensitivity was 92.2%, and the cumulative specificity was 61.9%.

3. Predictive capacity, sorted into three categories of sensitization potency

Within the GHS classification scheme, there is the possibility to refine the evaluation of skin sensitizers on the basis of their potency. Skin sensitizers can be assigned to subcategory 1A ("strong sensitizers") or to subcategory 1B ("other skin sensitizers") by using a weight of evidence approach on the basis of reliable and good quality data obtained from human cases or epidemiological studies and/or observations from appropriate studies in recognized and accepted animal tests. In the case of the LLNA, a threshold defined for the EC3 value was used to classify skin sensitizers into the two subcategories.

The performance of Cat. 1A/B classification was out of scope in this validation study, but Table 35 presents the predictive capacity for skin sensitization potential of the 27 test chemicals, as sorted into three categories of sensitization potency predicted by each of the three participating laboratories during the EpiSensA validation testing.

Table 35. Predictive capacity for skin sensitization potential of the 27 test chemicals, as sorted into three categories of sensitization potency.

	No. of chemicals	LION	KOSÉ	FDSC	Predictive capacity
1A	9	88.9%	88.9%	88.9%	88.9%
1B	9	88.9%	100.0%	100.0%	96.3%
No category	9	66.7%	66.7%	55.6%	63.0%

4. Comparison to the results of lead laboratory

The assessment of the predictive capacity forms only a secondary goal of the present validation study, not least since the sample size was determined for the purpose of satisfying the primary study goal and does not allow for robust conclusions to be drawn on the predictive capacity of the EpiSensA. Therefore, the predictive capacity during the validation study was compared to the results of lead laboratory.

The results of this validation study show that, when compared to LLNA test results, EpiSensA exhibits a sensitivity of 92.6%, a specificity of 63.0%, an accuracy of 82.7%, and a balanced accuracy of 77.8%. These results are comparable to the predictive performance based on 136 chemicals obtained by lead laboratory [a sensitivity of 88.1% (89/101), a specificity of 65.7% (23/35), an accuracy of 82.4% (112/136), and a balanced accuracy of 76.9%] (see Appendix 5).

When compared to human data, the results of this validation study show that EpiSensA exhibits a sensitivity of 92.2%, a specificity of 61.9%, an accuracy of 83.3%, and a balanced accuracy of 77.1%. These results are comparable to the predictive performance based on 80 chemicals obtained by lead laboratory [a sensitivity of 97.9% (46/47), a specificity of 48.5% (16/33), an accuracy of 77.5% (62/80), and a balanced accuracy of 73.2%] (see Appendix 5).

5. Test chemicals with results that were false negatives

Lauryl gallate

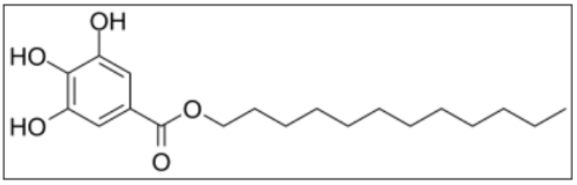
Although the LLNA results classify lauryl gallate as a sensitizer, it was predicted to be a non-sensitizer at all three participating laboratories (Table 36). Lauryl gallate did not show cytotoxicity at maximum soluble concentration (25 w/v%), so all participating laboratories performed main study at the highest three concentrations (6.25, 12.5, and 25 w/v%) except for 2nd experiment of FDSC. In contrast, the historical data showed that lauryl gallate slightly induced cytotoxicity and gene expressions to around the criteria only at the lower concentration (3.13 w/v%, Figure 28), which might be affected by AOO (penetration enhancing effect). Furthermore, lauryl gallate is a pre-hapten, and when oxidized, pre-haptens transform into reactive sensitizers. In EpiSensA, the RhE models are reported to show a metabolic capacity like that of the animal or human skin, suggesting that pre/pro-haptens might be correctly evaluated using the RhE model. On the other hand, the exposure time of 6 h in EpiSensA is shorter than that in an animal experiment such as the GPMT and the LLNA. If lauryl gallate needs more than 6 h to be sufficiently oxidized after penetrating a stratum corneum, it might be difficult to

accurately predict its sensitization potential through EpiSensA. Importantly, it is confirmed by lead laboratory that the fold-induction of *ATF3* and *IL-8* exceeded the respective cut-off values when 6 h exposure of lauryl gallate was followed by wash-out and post-incubation for 18 h (Figure 29). Therefore, it is suggested that EpiSensA would predict lauryl gallate as negative because of both high lipophilicity (logKow=6.21, suggesting low absorption at higher concentration) and autoxidation. Nevertheless, EpiSensA can test lipophilic chemicals by AOO adequately, and high sensitivity was confirmed (83% at 52 lipophilic sensitizers with logKow>3.5). In addition, EpiSensA can detect pre/pro-haptens with 97% (36/37) sensitivity.

From the point of view of combination with other assays, lauryl gallate is predicted to be a sensitizer by other in vitro methods such as the DPRA, KeratinoSens^{TM}, and h-CLAT (Appendix 12).

Table 36. Chemical structure of lauryl gallate, and overview of its prediction at the three participating laboratories.

Lauryl gallate	LION	KOSÉ	FDSC
1st	N	N	N
2nd	N	N	P
3rd	N	N	N
prediction	N	N	N
solvent	AOO	AOO	AOO



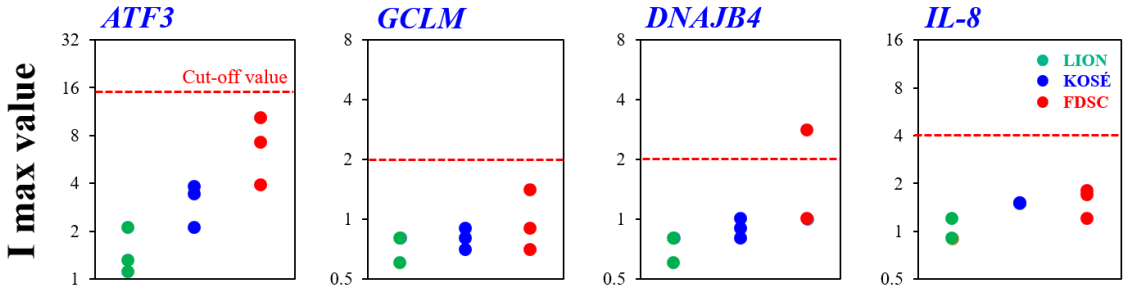


Figure 27. Distribution of the mean Imax values (n=3) for lauryl gallate.

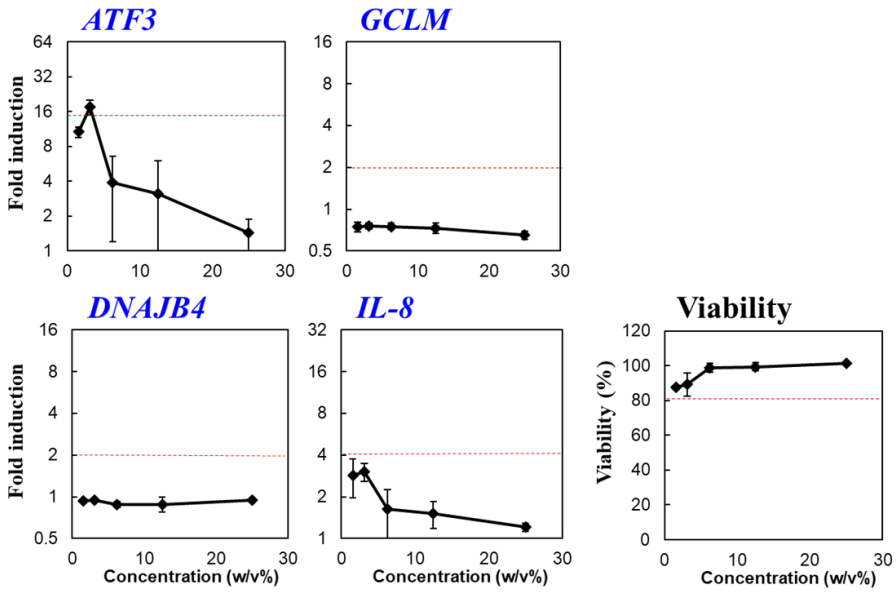


Figure 28. The historical data of lauryl gallate (n=3, bars: S.D.).

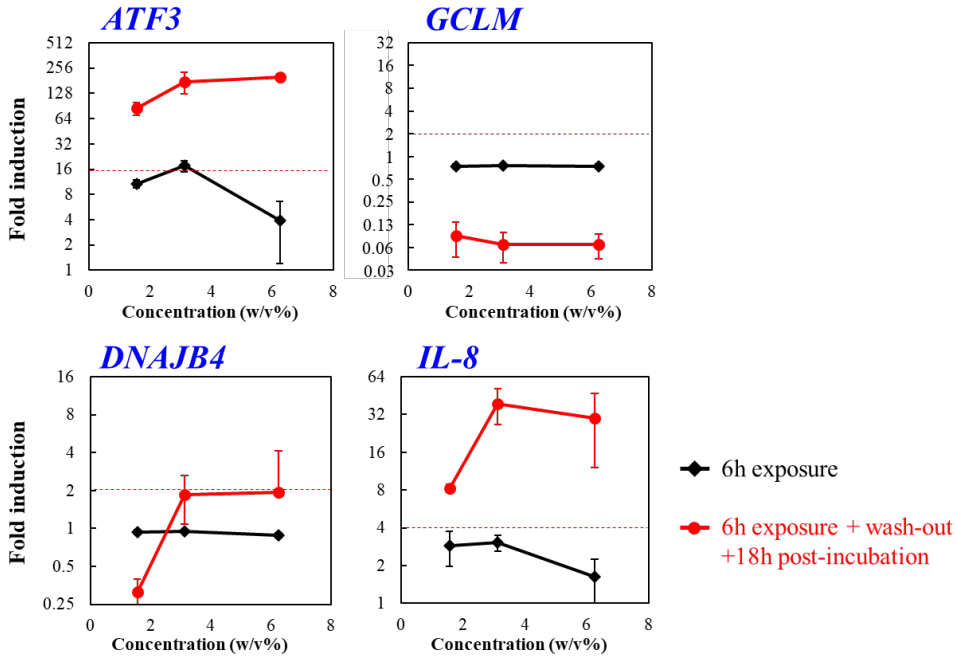


Figure 29. The post-incubation experiment of lauryl gallate (n=3, Baes: S.D.).

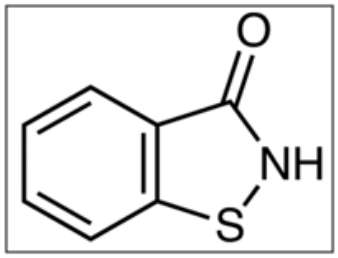
Benzisothiazolinone

Although the LLNA results classify benzisothiazolinone as a sensitizer, it was predicted to be a non-sensitizer by LION (Table 37). As shown in Figure 30, the Imax values for benzisothiazolinone were close to the cut-off values, and as such, it was

predicted to be a non-sensitizer by LION. For this reason, it might be difficult to obtain accurate results repeatedly, with a strong possibility of non-concordant results occurring across multiple participating laboratories. From the point of view of combination with other assays, benzisothiazolinone is predicted to be a sensitizer by other in vitro methods such as the DPRA, KeratinoSens^{TM}, and h-CLAT (Appendix 12).

Table 37. Chemical structure of lauryl gallate, and overview of its prediction obtained by the three participating laboratories.

Benzisothiazolinone	LION	KOSÉ	FDSC
1st	P	P	P
2nd	N	P	P
3rd	N	P	P
prediction	N	P	P
solvent	AOO	AOO	AOO



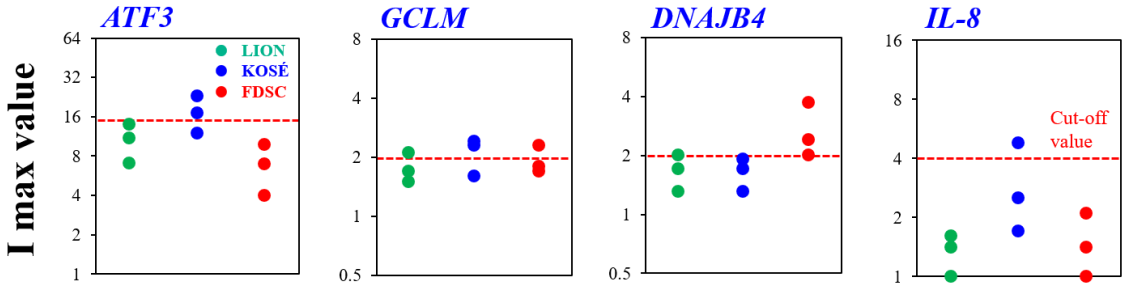


Figure 30. Distribution of the mean Imax values (n=3) for benzisothiazolinone.

6. Test chemicals with results that were false positives

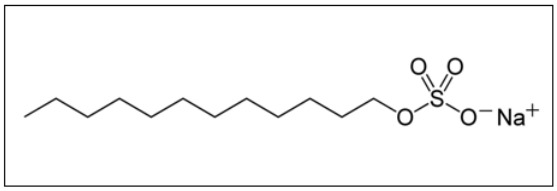
Sodium lauryl sulfate

SLS is an irritant and surfactant, and it has been reported to give false positive in LLNA. One potential reason for SLS giving false positive results in LLNA could be that SLS causes lymph node cell proliferation in mice through a non-immune mechanism involving the stimulation of cytokine production in the epidermis (Loveless *et al.*, 1996). In terms of EpiSensA, it has been reported that SLS induces the expression of IL-8 mRNA in keratinocytes (Torma *et al.*, 2006), while some irritating chemicals (including surfactants) can induce IL-8 expression in RhE (Saito *et al.*, 2017). Therefore, non-specific expression might occur in the EpiSensA in a similar manner to LLNA following an exposure to SLS. For this reason, SLS was predicted to be a sensitizer by the three

participating laboratories (Table 38). Here, not all irritating chemicals or surfactants induce non-specific *IL-8* expression. For example, erucamide and tween 80 were non-sensitizing surfactants and negative in EpiSensA (Imax value of *IL-8* was 1.4 and 2.2, respectively, from historical data). In addition, some skin irritating non-sensitizers (e.g., 1-bromobutane, lactic acid) were also judged as negative (Appendix 5). Here, SLS is predicted to be a non-sensitizer by DPRA, KeratinoSens™ and h-CLAT (Appendix 12).

Table 38. Chemical structure of sodium lauryl sulfate, and overview of its prediction at the three participating laboratories.

Sodium lauryl sulfate	LION	KOSÉ	FDSC
1st	P	P	P
2nd	P	P	P
3rd	P	P	P
prediction	P	P	P
solvent	AOO	AOO	AOO



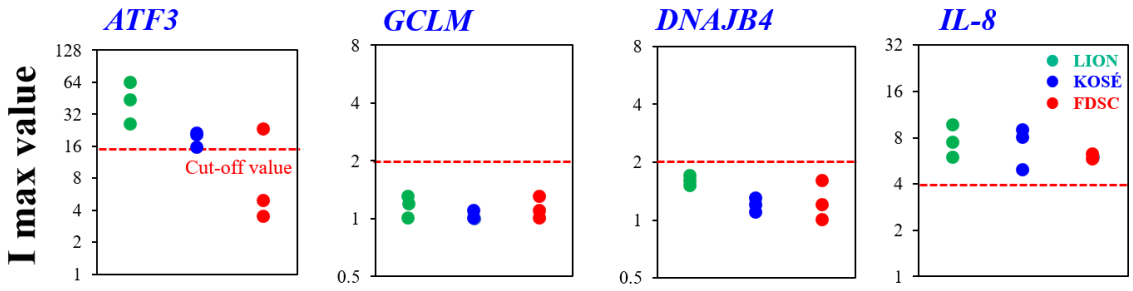


Figure 31. Distribution of the mean Imax values (n=3) for sodium lauryl sulfate.

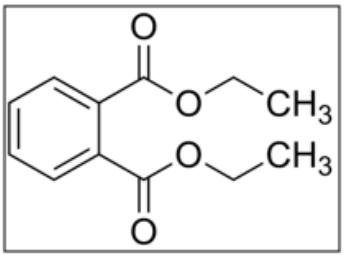
Diethyl phthalate

Although the LLNA results classify diethyl phthalate as a non-sensitizer, it was predicted to be a sensitizer by all three participating laboratories. Some phthalate diesters are known to be endocrine disrupting chemicals, and several epidemiological studies have associated these endocrine disrupting chemicals with reproductive effects. In terms of diethyl phthalate, it has been reported to induce testicular dysfunction in mice by activating both the transcription factor NFκB and Nrf2 (Mondal *et al.*, 2019). In addition, dibutyl phthalate (which is similar to diethyl phthalate) has been reported to activate the NLRP3 inflammasome through the $P2X_7$ receptor in HepG2 cells; an immortalized cell line deriving from a human hepatocellular carcinoma (Ni *et al.*, 2016). As described in the "Test Definition" section, NFκB regulates the induction of *ATF3*, Nrf2 regulates the induction of *GCLM* and *DNAJB4*, and $P2X_7$ regulates the induction of *ATF3* and *IL-8*.

Therefore, a non-specific expression of the aforementioned marker genes might occur as a result of an exposure to diethyl phthalate. For this reason, diethyl phthalate was predicted to be a sensitizer by all three participating laboratories (Table 39). Here, diethyl phthalate is predicted to be a sensitizer by h-CLAT and a non-sensitizer by DPRA and KeratinoSens™ (Appendix 12).

Table 39. Chemical structure of diethyl phthalate, and overview of its prediction at the three participating laboratories.

Diethylphthalate	LION	KOSÉ	FDSC
1st	P	P	P
2nd	P	P	P
3rd	P	P	P
prediction	P	P	P
solvent	AOO	AOO	AOO



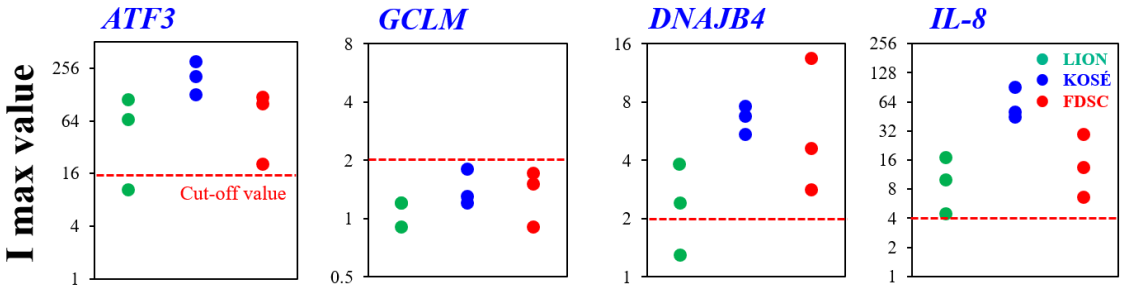


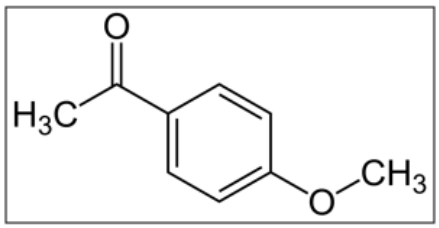
Figure 32. Distribution of the mean Imax values (n=3) for diethyl phthalate.

Acetanisole

Although the LLNA results classify acetanisole as a non-sensitizer, it was predicted to be a sensitizer by LION and FDSC (Table 40). At LION, the fold induction of the *ATF3* and *IL-8* expressions exceeded the respective cut-off values at the highest tested concentration, with the exception of doing so also at the lowest concentration (0.196%) at FDSC. In EpiSensA, it is extremely rare to witness a case where the excess of the cut-off value is achieved only at the lower tested concentration. In addition, as shown below in Table 40, the fold induction demonstrated a borderline induction for both the LION and the FDSC assessments. Based on the above, it is possible that a non-specific expression could occur as a result of an exposure to acetanisole, but the reason for this induction remains unclear. Here, acetanisole is predicted to be a sensitizer by KeratinoSens™ and a non-sensitizer by DPRA and h-CLAT (Appendix 12).

Table 40. Chemical structure of acetanisole, and overview of its prediction at the three participating laboratories.

Acetanisole	LION	KOSÉ	FDSC
1st	P	N	P
2nd	-	-	-
3rd	-	-	-
prediction	P	N	P
solvent	AOO	AOO	AOO



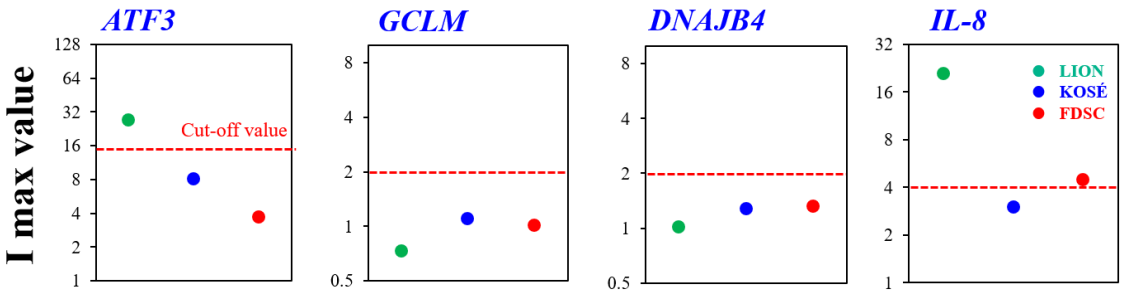


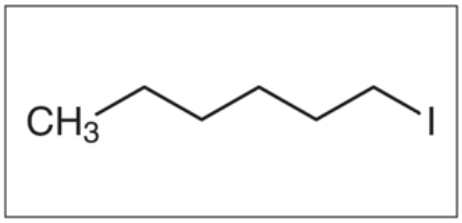
Figure 33. Distribution of the mean Imax values (n=3) for acetanisole.

1-Iodohexane

Although 1-iodohexane was selected as non-sensitizer in the validation study, it was predicted to be a sensitizer by the three participating laboratories (Table 41). It is likely that α-haloalkanes (such as 1-iodohexane) react with proteins via a nucleophilic substitution by a protein amino group on the α-carbon, with the corresponding displacement of a halide ion (Gerberick *et al.*, 2005). Moreover, it has been reported that 1-iodohexane forms covalent adducts with peptides (Natsch *et al.*, 2013), and that 1-iodohexane can dose-dependently induce the stimulation index in the LLNA by 0.9, 1.2 and 2.5 at 10%, 25% and 50%, respectively (Gerberick *et al.*, 2005). Moreover, 1-iodohexane is also predicted to be a sensitizer by other *in vitro* methods (such as the DPRA, KeratinoSens^{TM}, and h-CLAT) (Appendix 12). These results suggest that 1-iodohexane may truly possess skin sensitizing potential. Therefore, the positive results obtained by EpiSensA could be true positives.

Table 41. Chemical structure of 1-iodohexane, and overview of its prediction at the three participating laboratories.

1-Iodohexane	LION	KOSÉ	FDSC
1st	P	P	P
2nd	-	-	-
3rd	-	-	-
prediction	P	P	P
solvent	AOO	AOO	AOO



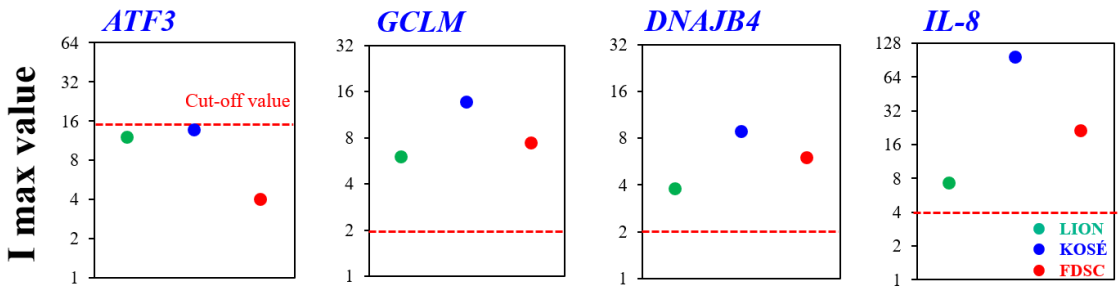


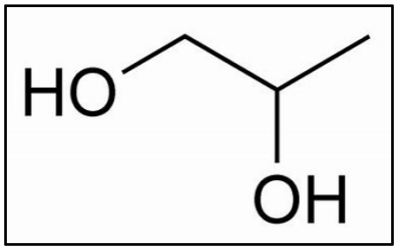
Figure 34. Distribution of the mean Imax values (n=3) for 1-iodohexane.

Propylene glycol

Although the LLNA results classify propylene glycol as a non-sensitizer, it was predicted to be a sensitizer by KOSÉ and FDSC (Table 42). Considering the hydrophilicity of propylene glycol, one potential reason could be its high osmotic stress. Notably, it has been reported that osmotic stress can induce the phosphorylation of the eukaryotic initiation factor 2 alpha (eIF2α) (Bevilacqua *et al.*, 2010); subsequently, the phosphorylation of eIF2α has been associated with increased expression of the *ATF3* (Cai & Brooks, 2011). Moreover, the fold induction of the *ATF3* expression has shown a borderline induction taking place only at the highest concentration in both the KOSÉ and FDSC experiments. Therefore, a non-specific expression of *ATF3* might occur as a result of an exposure to hydrophilic and low-cytotoxicity chemicals. For this reason, propylene glycol was predicted to be a sensitizer by KOSÉ and FDSC. Here, not all hydrophilic and low cytotoxic chemicals result in false-positive. These chemicals may tend to induce a low *ATF3* expression, but the induction usually lower than the cut-off value. For example, glucose was negative in EpiSensA (Imax value of *ATF3* was 6.5, from historical data). Here, propylene glycol is predicted to be a non-sensitizer by DPRA, KeratinoSens™, and h-CLAT (Appendix 12).

Table 42. Chemical structure of propylene glycol, and overview of its prediction at the three participating laboratories.

Propylene glycol	LION	KOSÉ	FDSC
1st	N	P	P
2nd	-	-	-
3rd	-	-	-
prediction	N	P	P
solvent	DW	DW	DW



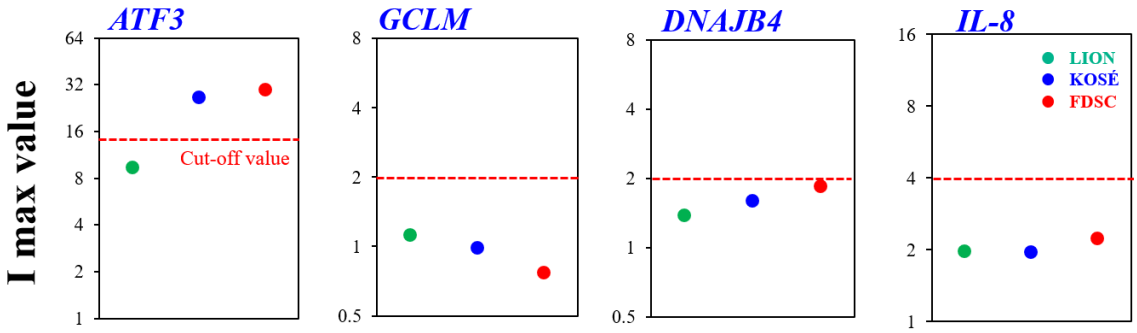


Figure 35. Distribution of the mean Imax values (n=3) for propylene glycol.

7. Conclusion

The VMT considered EpiSensA to be an *in vitro* test method which has shown potential to be used in combination with other information in defined approaches and integrated testing strategy (ITS) for skin sensitization hazard assessment. However, since the sample size is limited, which was defined for the assessment of the within and between laboratory reproducibility, robust conclusion on the predictive capacity should not be drawn. It is important that predictive capacity is assessed based on a much larger dataset.

Applicability domain

1. Technical limitations

1-1. Solubility of test chemicals

The EpiSensA test method as well as all other skin sensitization test methods, require the preparation of a stable test chemical solution or suspension with an appropriate vehicle. The SOP of EpiSensA contains provisions for the use of three different solvents (AOO, DW, 50% EtOH). The assay was performed using the vehicle that permitted the test chemical to be dissolved or allowed us to make a stable suspension at the highest concentration. If the chemical is not soluble or does not form a stable dispersion at 0.0122% in any vehicle, the chemical is not applicable for testing with EpiSensA. However, another vehicle could be employed if sufficient scientific rationale can be provided.

2. Summary of the SOP revision

The summary of revision history is provided in Table 43. The EpiSensA protocol was proven to be generally robust for the purposes of this study, with only minor clarifications and modifications being made to its SOP; these changes focused on specific elements of the procedure and the data interpretation in order to minimize the sources of variability.

The amendments introduced were largely aiming to resolve ambiguities and minor omissions in the original SOP, as well as to improve clarity and consistency of the data generation and interpretation. The VMT believes that the procedural clarifications to the EpiSensA original SOP, the supporting documents and the current study findings adequately demonstrate the status of the method's development and optimization, its mechanistic basis, its intended use and, the regulatory relevance of the EpiSensA.

Based on the comments from peer review panels, several modifications were included in the SOP version 2.6.

Table 43. Summary of the revision history of the EpiSensA SOP.

Revision	Date	Comments
Version 1.1	27th Jun. 2018	- Original SOP created by lead laboratory.
Version 2.1	1st Oct. 2018	- Positive control was changed from 6.25% w/v BADGE to 0.78% w/v clotrimazole and 0.10% w/v 4-nitrobenzylbromide. - The acceptance criterion for the effect on cell viability exerted by the vehicle control was changed from 80% to 90%. - The acceptance criterion regarding the GAPDH Ct value of the test chemicals was added. - The acceptance criteria for the vehicle control has been changed again.
Version 2.2	4th Sep. 2019	- The retesting criterion concerning a chemical with a steep dose-response curve was clarified. - The retesting criterion regarding the GAPDH Ct value was modified. - Solubilization procedure was slightly revised. - LDH assay protocol was modified in accordance with the kit manual.
Version 2.3	12th May 2020	- Note on topical application was revised. - cDNA synthesis method was slightly revised.
Version 2.4	23rd Sep. 2020	- Note on the liquid chemical exposure was added in order to avoid cross-contamination.
Version 2.5	27th Jul. 2021	- The topical application procedure was modified in order to prevent cross-contamination. - The note on topical application was revised.
Version 2.6	22nd Aug. 2022	- The purity for clotrimazole was added. - The tissue size of LabCyte EPI-MODEL 24 was added.

3. Limitations on applicability

2-1. Endogenous control gene

Chemicals that affect the expression of *GAPDH* might not be suitable for testing. When the *GAPDH* Ct value of each tested concentration is out of *GAPDH* Ct value of the corresponding vehicle control +/- 1, then the result at that concentration should not be considered. However, by using other endogenous control genes, these chemicals could be tested if sufficient scientific rationale can be provided.

2-2. Restrictions on cell viability assays

In the test method, the assays used for quantifying cell viability are the LDH assay. In addition, the MTT assay is also acceptable only if an inhibition of LDH reaction by a test chemical is estimated. Therefore, the test method might have a limitation for chemicals that directly interfere with both LDH and MTT.

2-3. Restriction on RNA extraction

Chemicals that either affect RNA itself (e.g., by causing RNA degradation) or directly interfere with the RNA isolation system, might not be suitable for testing.

2-4. Inherent limitation of a threshold-based prediction model

Not only EpiSensA but also other assays using a threshold-based prediction model have an inherent limitation in the detection of some weak sensitizers. The gene expressions induced by very weak sensitizers may remain just below the thresholds set for a positive result, and the chemicals may not be correctly classified.

2-5. Pre/pro-haptens

Lauryl gallate was judged as negative at three participating laboratories. However, it was confirmed that propyl gallate was correctly judged as positive at lead laboratories (Appendix 5). In addition, EpiSensA can detect pre/pro-haptens with 97.3% (36/37) sensitivity (Appendix 5). From these results, pre/pro-haptens are considered to be in domain of EpiSensA.

VMT overall conclusions and recommendations

1. Overall conclusions

The objective of this validation study was to assess the transferability, the within-laboratory reproducibility, and the between-laboratory reproducibility of EpiSensA. The study was undertaken by using coded test chemicals, enabling the VMT to assess the scientific validity and appropriateness of EpiSensA in anticipation of the development and issue of an OECD test method guideline for its use.

The VMT has concluded that the results of this validation study show that EpiSensA satisfies the requirements of modules 1 to 4 of the ECVAM Modular Approach to Validation (namely, for test definition, within-laboratory reproducibility, between-laboratory reproducibility, and transferability). In addition, these results contribute to module 5 (referring to the assay's predictive capacity) and module 6 (referring to the assay's applicability domain), and they are comparable to other validated test methods for the assessment of the skin sensitization potential of chemicals.

The main conclusions of the VMT in relation to each module are set out in the table below:

Module		Summary & Conclusions
1	Test definition	Both the existing body of evidence (original submission to JaCVAM including scientific publications) and the current study findings adequately demonstrate the intended purpose, the need, the status of development, and the mechanistic basis of the EpiSensA test method. An improved, well-detailed and robust SOP is available.
2	Within-laboratory reproducibility	The overall within-laboratory reproducibility was considered as acceptable for the proposed use of the EpiSensA (i.e., as part of an integrated testing strategy).
3	Transferability	The test method was shown to be transferable between laboratories.

		Training and demonstration of competence in the conduct of the assay is considered important. In particular, the chemicals used during this study's transferability stage should be considered.
4	Between-laboratory reproducibility	The between-laboratory reproducibility was acceptable.
5	Predictive capacity	A complete evaluation of the predictive capacity of EpiSensA was not one of the goals of this study. Taking into account chemicals tested outside this validation study, the VMT notes that the predictive performance of the assay for identifying skin sensitizers also has proved promising.

Overall, the VMT concludes that the information generated in this validation study shows that the EpiSensA is a robust and reliable test method. Consequently:

- Information generated by the EpiSensA can already be used in a weight-of-evidence approach in order to support regulatory decision making; for example, to characterize equivocal responses in *in vivo* studies (e.g., conflicting results from multiple studies).

- For the purposes of some regulations (for example REACH in the EU) a positive EpiSensA result of a test material could be considered a skin sensitizer.

- The EpiSensA is suitable for further evaluation as a component of a toolbox or as an ITS toward the full replacement of the *in vivo* assays currently used for skin sensitization hazard identification.

2. Recommendations

- The predictive accuracy of the EpiSensA should be evaluated in terms of its contribution to an integrated testing strategy for the full replacement of current *in vivo* hazard identification assays.

- Since the outcome of this validation study is consistent with the results obtained with the chemicals previously tested, and the revision of the SOP was solely for clarification on procedure, this validation study does not affect the historical results. Therefore, the existing/historical results should be taken into account for future formal evaluations on the predictive capacity.

Acknowledgement

This study received a research grant from the Japanese Society for Alternative to Animal Experiments, a grant-in-aid from the Japan Agency for Medical Research and Development (AMED) under Grant Number JP20mk0101131 and the Long-range Research Initiative by Japan Chemical Industry Association.

References

Amber B Daniel, Judy Strickland, David Allen, Silvia Casati, Valérie Zuang, João Barroso, Maurice Whelan, M J Régimbald-Krnel, Hajime Kojima, Akiyoshi Nishikawa, Hye-Kyung Park, Jong Kwon Lee, Tae Sung Kim, Isabella Delgado, Ludmila Rios, Ying Yang, Gangli Wang, Nicole Kleinstreuert, (2018). International regulatory requirements for skin sensitization testing, 95:52-65. doi: 10.1016/j.yrtph.2018.03.003.

Andres, E., Barry, M., Hundt, A., Dini, C., Corsini, E., Gibbs, S., Roggen, E.L., Ferret, P.J. (2017). Preliminary performance data of the RHE/IL-18 assay performed on SkinEthicTM RHE for the identification of contact sensitizers. International Journal of Cosmetic Science, 39(2), 121-132. doi: 10.1111/ics.12355.

Aptula, A. O., Roberts, D. W., Pease, C. K. (2007). Haptens, prohaptens and prehaptens, or electrophiles and proelectrophiles. Contact Dermatitis, 56(1), 54-56. doi: 10.1111/j.1600-0536.2007.00944.x

Bevilacqua, E., Wang, X., Majumder, M., Gaccioli, F., Yuan, C.L., Wang, Chuanping, Zhu, X., Jordan, L.E., Scheuner, D., Kaufman, R.J., Koromilas, A.E., Snider, M.D., Holcik, M., Hatzoglou, M. (2010). EIF2alpha phosphorylation tips the balance to apoptosis during osmotic stress. Journal of Biological Chemistry, 285(22), 17098-17111. doi: 10.1074/jbc.M110.109439.

Casati, S., Aeby, P., Kimber, I., Maxwell, G., Ovigne, J.M., Roggen, E., Rovida, C., Tosti, L., Basketter, D. (2009). Selection of chemicals for the development and evaluation of in vitro methods for skin sensitisation testing. Alternatives to Laboratory Animals: ATLA, 37(3), 305-312. doi: 10.1177/026119290903700313.

Cottrez, F., Boitel, E., Ourlin, J.C., Peiffer, J.L., Fabre, I., Henaoui, I.S., Mari, B., Vallauri, A., Paquet, A., Barbry, P., Auriault, C., Aeby, P., Groux, H. (2016). SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. Toxicology in Vitro: An International Journal Published in Association with BIBRA, 32, 248-260. doi: 10.1016/j.tiv.2016.01.007.

EU (2009). Regulation (EC) no 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. Official Journal of the European Union, L 342, 59-209.

Franklin, C.C., Backos, D.S., Mohar, I., White, C.C., Forman, H.J., Kavanagh, T.J. (2009). Structure, function, and post-translational regulation of the catalytic and modifier

subunits of glutamate cysteine ligase. Molecular Aspects of Medicine, 30(1-2), 86-98. doi: 10.1016/j.mam.2008.08.009.

Frank Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A. (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. Dermatitis: Contact, Atopic, Occupational, Drug, 16(4), 157-202.

Gibbs S, Corsini E, Spiekstra SW, Galbiati V, Fuchs HW, Degeorge G, Troese M, Hayden P, Deng W, Roggen E. An epidermal equivalent assay for identification and ranking potency of contact sensitizers. Toxicol Appl Pharmacol. 2013 Oct 15;272(2):529-41. doi: 10.1016/j.taap.2013.07.003.

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

Hans Törmä, H., Geijer, S., Gester, T., Alpholm, K., Berne, B., Lindberg, M. (2006). Variations in the mRNA expression of inflammatory mediators, markers of differentiation and lipid-metabolizing enzymes caused by sodium lauryl sulphate in cultured human keratinocytes. Toxicology in Vitro: An International Journal Published in Association with BIBRA, 20(4), 472-479. doi: 10.1016/j.tiv.2005.09.011.

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V. (2004). A modular approach to the ECVAM principles on test validity. Alternatives to Laboratory Animals: ATLA, 32(5), 467-472. doi: 10.1177/026119290403200503.

Hoffmann, E., Dittrich-Breiholz, O., Holtmann, H., Kracht, M. (2002). Multiple Control of interleukin-8 Gene Expression. Journal of Leukocyte Biology, 72(5), 847-855. doi: doi.

Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, J., Lalko, J.F., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D. (2018). Non-animal methods to predict skin sensitization (I): The Cosmetics Europe database<sup/>. Critical Reviews in Toxicology, 48(5), 344-358. doi: 10.1080/10408444.2018.1429385.

Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., Miyazawa, M. (2015). Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment:

A decision support system for quantitative weight of evidence and adaptive testing strategy. Archives of Toxicology, 89(12), 2355-2383. doi: 10.1007/s00204-015-1634-2.

Loveless, S.E., Ladics, G.S., Gerberick, G.F.., C A Ryan., D A Basketter., E W Scholes., R V House., J Hilton., R J Dearman., I Kimber. (1996). Further evaluation of the local lymph node assay in the final phase of an international collaborative trial, Toxicology.

Leonard, E.J., Skeel, A., Yoshimura, T., Noer, K., Kutvirt, S., Van Epps, D., 1990. Leukocyte specificity and binding of human neutrophil attractant/activation protein-1. JI 144, 1323-1330.

Lu, S.C. (2013). Glutathione synthesis. Biochimica et Biophysica Acta, 1830(5), 3143-3153. doi: 10.1016/j.bbagen.2012.09.008.

McKim, J.M. Jr., Keller, D.J. 3rd, Gorski, J.R. (2012). An in vitro method for detecting chemical sensitization using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. Cutaneous and Ocular Toxicology, 31(4), 292-305. doi: 10.3109/15569527.2012.667031.

Mizumachi, H., Sakuma, M., Ikezumi, M., Saito, K., Takeyoshi, M., Imai, N., Okutomi, H., Umetsu, A., Motohashi, H., Watanabe, M., Miyazawa, M. (2018). Transferability and within- and between-laboratory reproducibilities of EpiSensA for predicting skin sensitization potential in vitro: A ring study in three laboratories. Journal of Applied Toxicology: JAT, 38(9), 1233-1243. doi: 10.1002/jat.3634.

Mondal, S., Ghosh, S., Bhattacharya, S., Mukherjee, S. (2019). Chronic dietary administration of lower levels of diethyl phthalate induces murine testicular germ cell inflammation and sperm pathologies: Involvement of oxidative stress. Chemosphere, 229, 443-451.

Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., Kern, P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. Journal of Applied Toxicology: JAT, 33(11), 1337-1352. doi: 10.1002/jat.2868.

Jiahua Ni, J., Zhang, Z., Luo, X., Xiao, L., Wang, N. (2016). Plasticizer DBP activates NLRP3 inflammasome through the P2X7 receptor in HepG2 and L02 cells. Journal of Biochemical and Molecular Toxicology, 30(4), 178-185. doi: 10.1002/jbt.21776.

OECD (2012)10/PART1. The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. Part 1: Scientific evidence. Series on Testing and Assessment. ENV/JM. Mono, 168.

OECD. (1992). Guideline for Testing of Chemicals No. 406: Skin Sensitization, adopted 17th July.

OECD (1998), OECD Principles on Good Laboratory Practice, OECD SERIES ON PRINCIPLES OF GOOD LABORATORY PRACTICE AND COMPLIANCE MONITORING, No 1, Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/mc/chem(98)17&doclanguage=en

OECD. (2010). Guideline for Testing of Chemicals No. 429: Skin Sensitization: Local lymph node Assay, adopted 22nd July.

OECD. (2016a). Guidance document on the reporting of defined approaches to be used within integrated approaches to testing and assessment. Series on testing and Assessment No. 255. OECD Publishing, Paris.

OECD. (2016b). Guidance document on the reporting of defined approaches to be used within integrated approaches to testing and assessment. Series on testing and Assessment No. 256. OECD Publishing, Paris.

OECD. (2018a). KEY EVENT BASED TEST guideline 442D: In vitro SKIN SENSITISATION ASSAYS ADDRESSING THE AOP KEY EVENT ON keratinocyte ACTIVATION, adopted 25th June.

OECD. (2018b). KEY EVENT BASED TEST guideline 442e: In vitro SKIN SENSITISATION ASSAYS ADDRESSING THE KEY EVENT ON ACTIVATION OF dendritic CELLS ON THE ADVERSE OUTCOME PATHWAY FOR SKIN SENSITISATION, adopted 25th June.

OECD. (2019). Guidelines for the testing of chemicals: In Chemico Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins, adopted 18th June No. 442c.

OECD. (2021). Guideline No. 497, Guideline on Defined Approaches for Skin Sensitisation, adopted 14th June.

Oesch, F., Fabian, E., Landsiedel, R. (2018). Xenobiotica-metabolizing enzymes in the skin of rat, mouse, pig, guinea pig, man, and in human skin models. Archives of Toxicology, 92(8), 2411-2456. doi: 10.1007/s00204-018-2232-x.

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

Onami, K., Kimura, Y., Ito, Y., Yamauchi, T., Yamasaki, K., Aiba, S. (2014). Nonmetal haptens induce ATP release from keratinocytes through opening of pannexin

hemichannels by reactive oxygen species. The Journal of Investigative Dermatology, 134(7), 1951-1960. doi: 10.1038/jid.2014.93.

Qi Cai, Q., Brooks, H.L. (2011). Phosphorylation of eIF2α via the general control kinase, GCN2, modulates the ability of renal medullary cells to survive high urea stress. American Journal of Physiology. Renal Physiology, 301(6), F1202-F1207. doi: 10.1152/ajprenal.00272.2011.

Saito, K., Takenouchi, O., Nukada, Y., Miyazawa, M., Sakaguchi, H. (2017). An in vitro skin sensitization assay termed EpiSensA for broad sets of chemicals including lipophilic chemicals and pre/pro-haptens. Toxicology in Vitro: An International Journal Published in Association with BIBRA, 40, 11-25. doi: 10.1016/j.tiv.2016.12.005.

Thompson, M.R., Xu, D., Williams, B.R. (2009). ATF3 transcription factor and its emerging roles in immunity and cancer. Journal of Molecular Medicine, 87(11), 1053-1060. doi: 10.1007/s00109-009-0520-x.

Tokudome, Y., Katayanagi, M., Hashimoto, F. (2015). Esterase activity and intracellular localization in reconstructed human epidermal cultured skin models. Annals of Dermatology, 27(3), 269-274. doi: 10.5021/ad.2015.27.3.269.

Weber, F.C., Esser, P.R., Müller, T., Ganesan, J., Pellegatti, P., Simon, M.M., Zeiser, R., Idzko, M., Jakob, T., Martin, S.F. (2010). Lack of the purinergic receptor P2X(7) results in resistance to contact hypersensitivity. Journal of Experimental Medicine, 207(12), 2609-2619. doi: 10.1084/jem.20092489.

Weber, F.C., Németh, T., Csepregi, J.Z., Dudeck, A., Roers, A., Ozsvári, B., Oswald, E., Puskás, L.G., Jakob, T., Mócsai, A., Martin, S.F. (2015). Neutrophils are required for both the sensitization and elicitation phase of contact hypersensitivity. Journal of Experimental Medicine, 212(1), 15-22. doi: 10.1084/jem.20130062.