Report on a Validation Study of the IL-2 Luc Assay for Evaluating the Potential Immunotoxic Effects of Chemicals on T-Cells

Validation Management Team

1. Summary

The IL-2 luciferase reporter assay (IL-2 Luc assay) was developed as one of 3 luciferase reporter assays in the Multi-ImmunoTox assay (MITA), a high-throughput screening system that our group had developed to evaluate chemical immunotoxicity. Although our final long-term goal is to officially validate the MITA for within- and between- laboratory reproducibility and predictivity, in this study, we conducted the validation for IL-2 Luc assay as the initial step.

In the MITA, we used 3 stable lines of reporter cells transfected with luciferase genes under control of the IL-2, IFN-γ, IL-8, and IL-1β promoters: 2H4 cells derived from Jurkat cells containing stable luciferase green (SLG) regulated by the IL-2 promoter, stable luciferase orange (SLO) regulated by the IFN-γ promoter, and stable luciferase red (SLR) regulated by the GAPDH promoter; THP-G8 cells derived from THP-1 cells containing SLO regulated by the IL-8 promoter and SLR regulated by the GAPDH promoter; and THP-G1b cells derived from THP-1 cells containing SLG regulated by the IL-1β promoter and SLR regulated by the GAPDH promoter. We selected these 4 cytokines because IL-2 and IFN-γ are primarily produced by T cells (a type of adaptive immune cells), whereas IL-8 and IL-1β are primarily produced by monocytes and dendritic cells (types of innate immune cells).

Using these 3 cell lines, the MITA can evaluate the effects of chemicals on the IL-2 and IFN-γ luciferase activity of 2H4 cells stimulated with phorbol 12-myristate 13-acetate (PMA) and ionomycin (Io), those on the IL-1β and IL-8 luciferase activity of THP-G1b and THP-G8 cells, respectively, stimulated by lipopolysaccharide (LPS).

In the validation study of the IL-2 Luc assay, the preliminary test trial, Phase 0, was performed by the participating laboratories following explicit explanations of the Multi-ImmunoTox Assay protocol Ver. 008.1E proposed by the lead laboratory, Tohoku University. Three laboratories participated in the Phase 0 study of the IL-2 Luc assay using 5 open labeled chemicals (2-aminoantracene, citral, chloroquine, dexamethasone and methyl mercuric chloride), in which they conducted 1 set composed of 3

experiments for each chemical. Most response patterns for the 5 chemicals were similar among the 3 laboratories, except for 2 early experiments conducted by the naïve laboratory. Based on these results, the Validation Management Team (VMT) judged that technical and protocol transfer of the IL-2 Luc assay is acceptable.

In the Phase I study, a total of 5 coded chemicals (4 T cell targeting and 1 non-T cell targeting) were evaluated by 3 experimental sets based on the Multi-ImmunoTox Assay protocol Ver. 011E made by the lead laboratory, Tohoku University. The average within-laboratory reproducibility was 86.7% (13/15). The between-laboratory reproducibility was 80.0% (4/5). The average predictivity was 93.3% (14/15).

In the Phase II study, between-laboratory reproducibility and predictivity using a total of 20 coded chemicals (13 T cell targeting, 6 non-T cell targeting, and 1 undetermined) were evaluated by 1 experiment set based on the Multi-ImmunoTox Assay protocol Ver. 009.1E. The between-laboratory reproducibility was 80% (16/20) and the average predictivity was 70.2% (40/57).

In the combined results of the Phase I and II studies, the average within-laboratory reproducibility was 86.7% (13/15). The between laboratory reproducibility was 80% (20/25). The average predictivity was 75.0% (54/72).

Although the within- and between-laboratory reproducibilities could satisfy the acceptance criteria for the validation study, the predictivity was below 80%. We considered several possible reasons for this unsatisfactory predictivity.

Since the 2H4 cell line used in the IL-2 Luc assay is derived from Jurkat cells, the IL-2 Luc assay cannot evaluate immunotoxic effects of immunosuppressive compounds whose mode of action is the inhibition of DNA synthesis leading to myelotoxicity. Thus, these chemicals should be outside the defined applicability domain for the assay. To overcome this limit, the IL-2 Luc assay requires combination with assays capable of detecting myelotoxicity, such as the conventional 28-day repeat dose toxicity test or *in vitro* myelotoxicity tests (Pessina et al., 2003). In addition, chemicals that need metabolic activation or poor water soluble need to be outside the applicability domain.

Even though these applicability domains are taken into consideration, the IL-2 Luc assay alone cannot cover all the effects of chemicals on human immune system. Therefore, it is indispensable to develop other *in vitro* systems to detect the effects of chemicals on different aspects of immune response. By accumulating and combining various approaches to detect chemical immunotoxicity, the *in vitro* assays can cover the effects of chemicals on the broad range of human immune system. The IL-2 Luc assay can be the first step.

2. Objective of the study

The objective of the present validation study was to determine the usefulness and limitations of the IL-2 Luc assay in MITA as a non-animal screening method to detect and assess the immunotoxicity of chemicals.

The specific objectives of the study were to establish:

1) “Transferability”, i.e., the extent to which a laboratory can adapt and easily implement the IL-2 reporter assay;

2) “Between or inter-laboratory reproducibility”, i.e., the extent to which results agree among different laboratories;

3) "Within or intra-laboratory reproducibility", i.e., the extent to which results agree in the same laboratory; and

4) “Predictivity”, i.e., the extent to which the *in vitro* results agree with the known immunological profiles of the chemicals.

3. Background

3-1. What is immunotoxicity?

A well-functioning immune system is essential for maintaining the integrity of an organism. Immune dysregulation can have serious adverse health consequences, ranging from reduced resistance to infection and neoplasia to allergic and autoimmune conditions. Environmental contaminants, food additives, and drugs can target the immune system, resulting in immune dysregulation. Accordingly, the potential for immunotoxicity, which is defined as the toxicological effects of xenobiotics on the function of the immune system, has raised serious concerns from the public as well as regulatory agencies. Currently, the assessment of chemical immunotoxicity relies mainly on animal models and assays that characterize immunosuppression and sensitization. However, animal studies have many drawbacks, such as high cost, ethical concerns, and questionable relevance to risk assessment for humans.

3-2. The current status of *in vitro* approaches to detect immunotoxicants

Now the worldwide vision is promoting alternative testing methods and assessment strategies to reduce the use of laboratory animals and, if possible, replace animals used in scientific studies (Adler et al., 2011). The workshop hosted by the European Centre for the Validation of Alternative Methods (ECVAM) in 2003 focused on state-of-the-art *in vitro* systems for evaluating immunotoxicity (Galbiati et al., 2010; Gennari et al., 2005; Lankveld et al., 2010). In the ECVAM workshop, a tiered approach was proposed. Since useful information can be obtained from regular 28-day general toxicity tests, pre-screening for direct immunotoxicity would begin with the evaluation of myelotoxicity in the proposed tiered approach (Corsini and Roggen, 2017). Compounds that are capable of damaging or destroying bone marrow will most likely have immunotoxic effects. If compounds are not potentially myelotoxic, they are tested for leukotoxicity. Compounds are then tested for immunotoxicity using various approaches such as the human whole-blood cytokine release assay (HWBCRA), lymphocyte proliferation assay, mixed lymphocyte reaction, NK cell assay, T cell–dependent antibody response, dendritic cell maturation assay, and fluorescent cell chip (FCP) assay. Among these assays, the HWBCRA has undergone formal pre-validation, although other techniques are being examined or have been examined in a rigorous pre-validation effort by the ECVAM and other groups. (Fig. 1) However, these assays require fresh rodent or human immune cells, in conflict with animal protection goals. The need for primary cells may decrease reproducibility and makes the assay unsuitable for high-throughput approaches

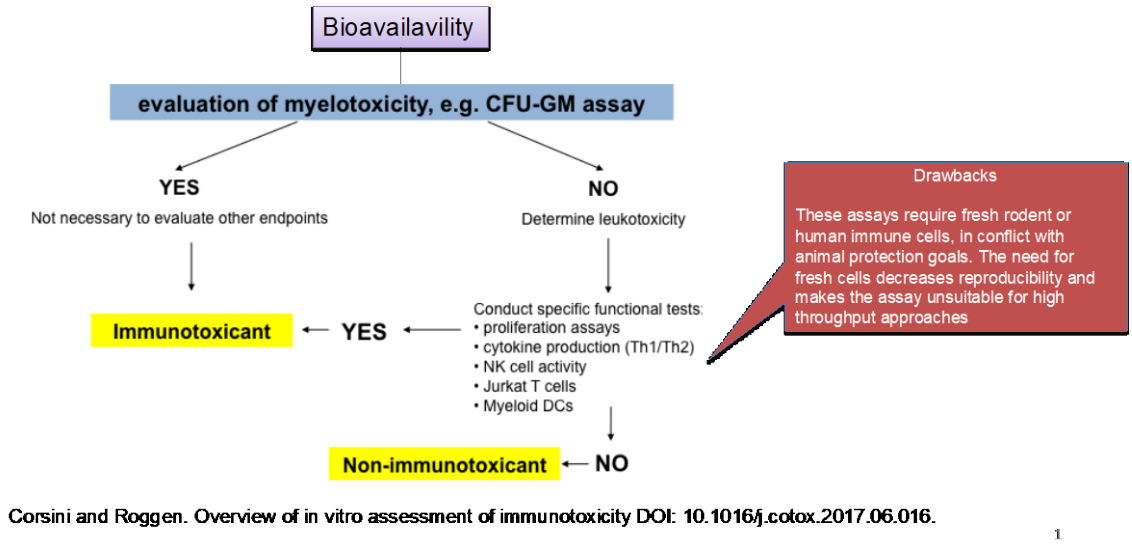


Fig. 1. Decision tree approach for *in vitro* assessment of chemical-induced immunosuppression.

3-3. *In vitro* immunotoxicity tests in principle should evaluate effects on both innate and acquired immunity

The immune system comprises innate and adaptive immunity (Fig. 2). Both arms of the immune response function differently and are driven by different populations of cells. In innate immunity, pathogens are recognized through various pattern recognition molecules, such as C-type lectin receptors, toll-like receptors, nod-like receptors, and retinoic acid–inducible gene-I (RIG-I)-like receptors. In addition, a variety of different cells are involved in this type of response, including neutrophils and other types of granulocytes, macrophages, natural killer (NK) cells, innate lymphoid cells, and mast cells. Adaptive immune responses involve specific antigen receptors encoded by rearranged genes, and T cells and B cells play critical roles in these responses.

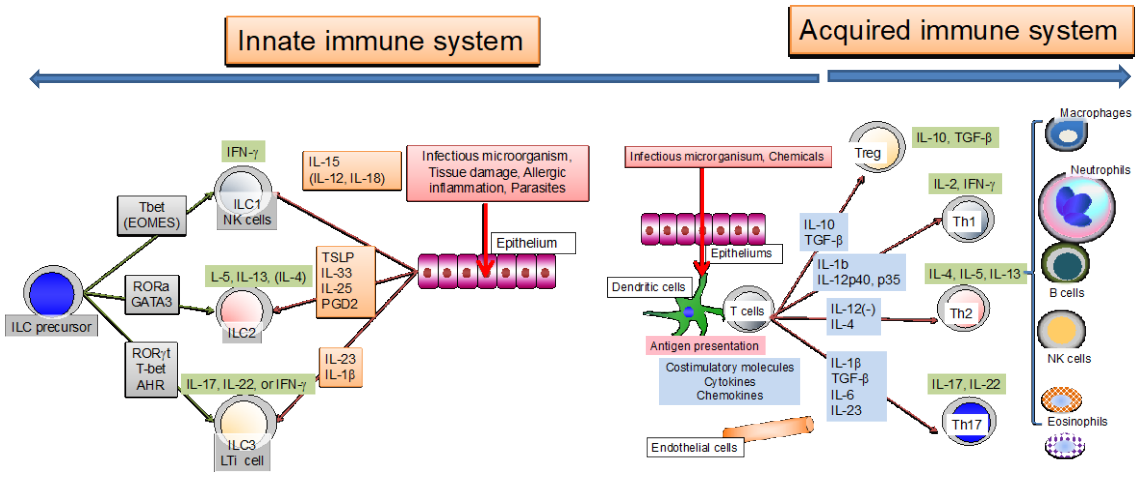


Fig2. Schematic representation of the innate immune system and acquired immune system.

Macrophages and dendritic cells (DCs), which act as antigen-presenting cells (APCs), link the innate and adaptive immune responses because they can present antigens to T lymphocytes in the context of major histocompatibility complex (MHC) class I or II molecules and stimulate their proliferation and effector functions after being stimulated via pathogen recognition receptors (Fig. 3). To induce optimal immune responses to various pathogens and minimize autoreactivity, innate and adaptive immune cells produce a vast array of cytokines, chemokines, and chemical mediators and present the molecules required for direct cell-cell interaction on their surface. A variety of intracellular signaling pathways also play roles in innate and adaptive immune responses.

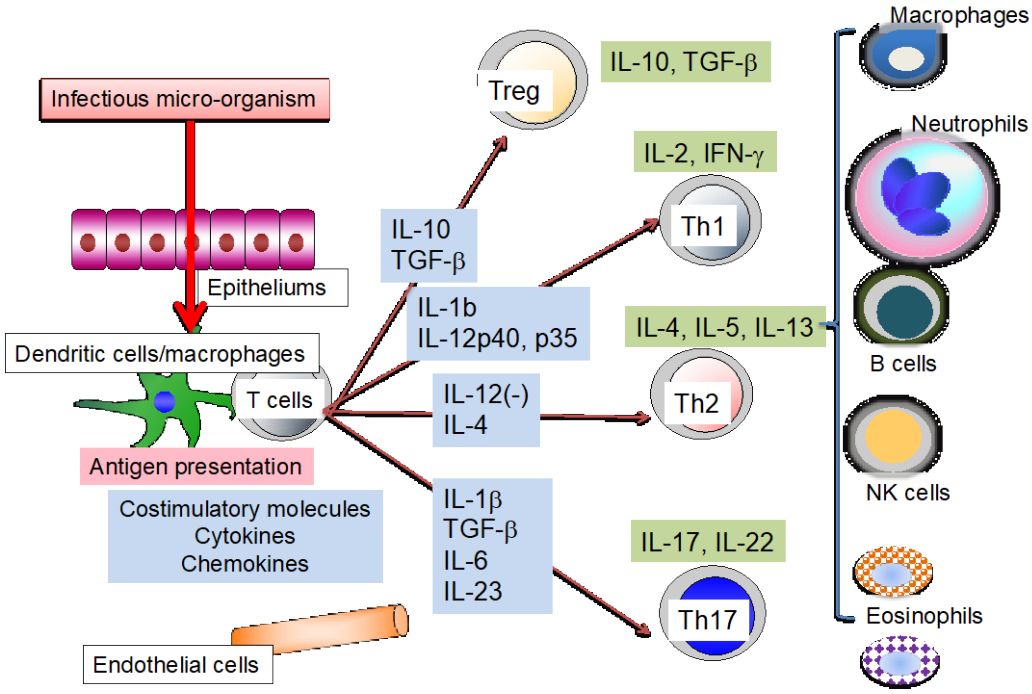


Fig. 3. Dendritic cells link the innate immune response to the acquired immune response.

Theoretically, chemicals can affect the immune system by targeting either the innate immune system or the acquired immune system (Fig. 2 and Fig. 3). Therefore, novel *in vitro* test methods are needed to adequately assess the immunotoxic effects of chemicals on both arms of immune system.

3-4. Mechanism for the induction of immunotoxicity by chemicals

Given the complexity of the immune system, it is unlikely that a single *in vitro* method will be able to detect all immunotoxicants. The mechanisms underlying the immunotoxicity of chemicals can be classified into 3 main categories: 1) killing of immune cells caused by bone marrow toxicity, 2) interference with general or immune-specific signaling leading to changes in the expression of cell surface molecules, cytokines or chemokine production, cell differentiation, and activation, and 3) binding to proteins forming complete antigens or exposing cryptic self-proteins (Fig. 4).

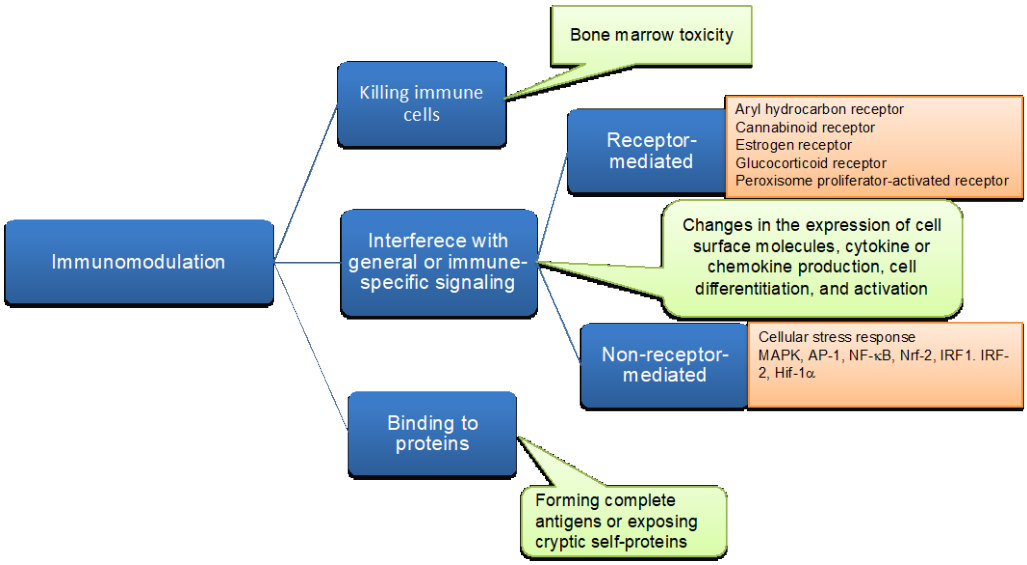


Fig. 4. Main mechanisms of immunotoxicity

Chemicals can interfere with immune-related cell signaling through receptor-mediated pathways using xenobiotic receptors such as the aryl hydrocarbon receptor (AhR), constitutive androstane receptor (CAR), or pregnane X receptor (PXR) (Elentner et al., 2018; Hidaka et al., 2017), cannabinoid receptor, estrogen receptor, glucocorticoid receptor or peroxisome proliferator-activated receptor or through non-receptor-mediated ways. Without specific receptors, it has been demonstrated that so-called cellular stress response can cause immunotoxicity (Fulda et al., 2010; Kultz, 2005). In essence, as long as stress stimulus does not cross a certain threshold, a cell can cope and survive by mounting an appropriate protective response. Conversely, the failure to activate or maintain a protective response (e.g., when the stressor is too strong) results in activation of stress signaling cascades that eventually activate cell death pathways. Depending on the type of stress and its severity, a cell's response can be manifold. However, most cellular protective responses induced by chemicals can be classified into one of several

categories, such as heat shock, unfolded protein, DNA damage, and oxidative stress responses, in addition to the response to danger signals (Gallucci and Matzinger, 2001). These responses are independent of the chemical species (Fig. 5). In addition, these cellular stress responses can affect immune function because they share the same cellular signaling pathways, e.g., MAP kinase, NF-κB, and mTOR, used by the immune response (Milisav, 2011). Indeed, although sensitizers that induce allergic contact hypersensitivity include numerous compounds with different molecular structures, it has become clear that their ability to sensitize is based simply on their reactivity to cysteine residues, which induces a response to oxidative stress (Sasaki and Aiba, 2007). Therefore, although it is assumed that there may be many chemicals with the potential to produce immunotoxicity, only a limited number of assay systems may be required to detect their effects.

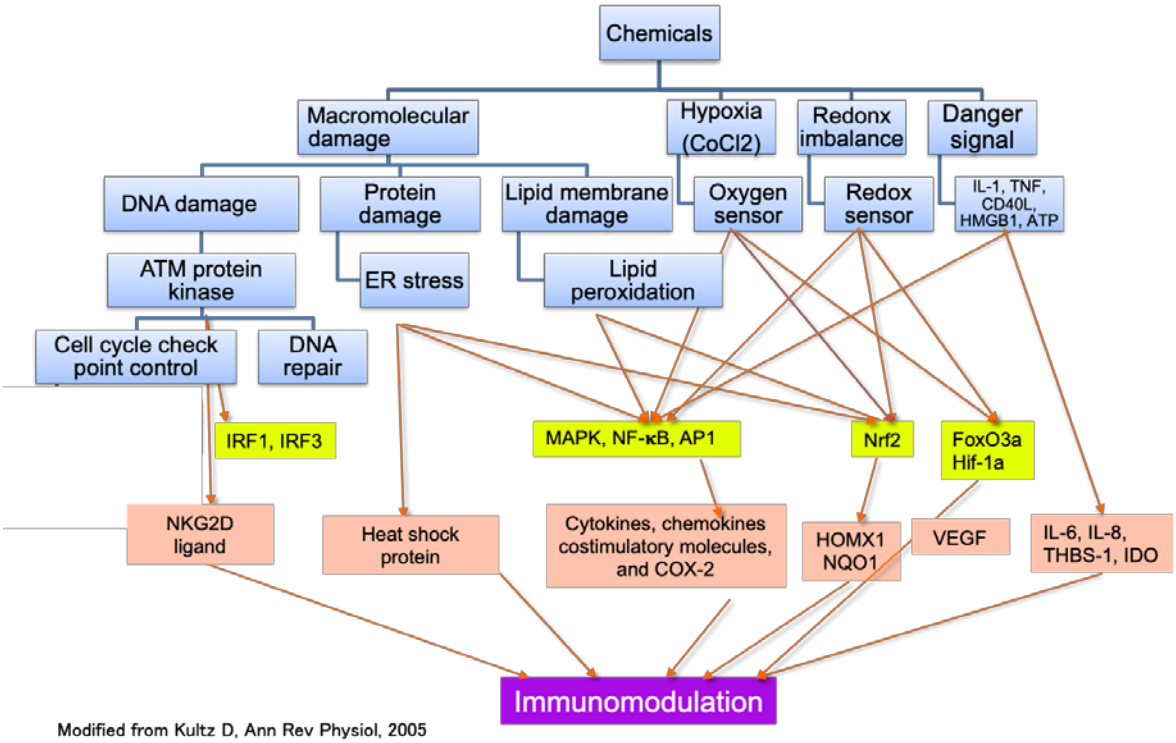


Fig. 5. Cellular stress response and danger signals.

3-5. Multi-ImmunoTox assay (MITA)

Our group developed a high-throughput screening system to evaluate chemical immunotoxicity. We first established 3 stable reporter cell lines transfected with luciferase genes under control of the IL-2, IFN-γ, IL-8, and IL-1β promoters: 2H4 cells derived from Jurkat cells containing stable luciferase green (SLG) regulated by the IL-2 promoter, stable luciferase orange (SLO) regulated by the IFN-γ promoter, and stable luciferase red (SLR) regulated by the GAPDH promoter (Saito et al., 2011); THP-G8 cells derived from THP-1 cells containing SLO regulated by the IL-8 promoter and SLR regulated by GAPDH promoter (Takahashi et al., 2011); and THP-G1b cells derived from THP-1 cells containing SLG regulated by the IL-1β promoter and SLR by the GAPDH promoter (Kimura et al., 2014). These 4 cytokines were selected because IL-2 and IFN-γ are primarily produced by T cells (adaptive immune cells), whereas IL-8 and IL-1β are primarily produced by monocytes and dendritic cells (innate immune cells). Using these 3 cell lines, we established the Multi-ImmunoTox assay (MITA). This assay identifies the effects of chemicals on the IL-2 and IFN-γ luciferase activity in 2H4 cells in the presence of the stimulants phorbol 12-myristate 13-acetate (PMA) and ionomycin (Io), and on the IL-1β and IL-8 luciferase activities in THP-G1b and THP-G8 cells, respectively, in the presence of the stimulant lipopolysaccharide (LPS) (Fig. 6).

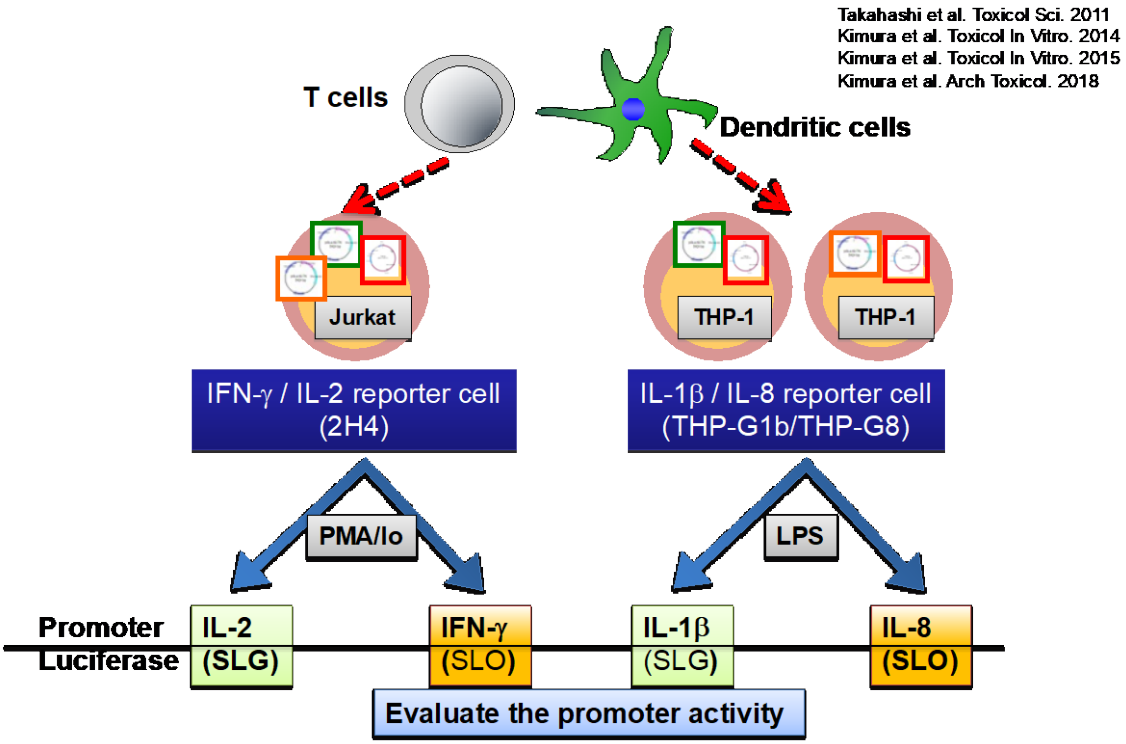


Fig. 6. The Muti-ImmunoTox assay (MITA)

3-6. The luciferase activities of the three MITA cell lines correspond with mRNA expression in the wild type cell lines or in human whole blood cells when stimulated with PMA/Io or LPS in the presence of 3 representative immunosuppressive drugs

After establishing the MITA, we first compared the effects of dexamethasone, cyclosporine, and tacrolimus on the 3 MITA cell lines with those on mRNA expression in the wild type cell lines or in human whole-blood cells stimulated with PMA/Io or LPS. The results confirmed that the MITA correctly reflects changes in mRNA expression in the mother cell lines and whole-blood cells (Kimura et al., 2014).

3-7. The MITA can evaluate the immunotoxicity profiles of well-known immunosuppressive drugs

We next evaluated the performance of the MITA by examining immunosuppressive or immunomodulatory drugs with well-known clinical effects on the human immune system (Kimura et al., 2014). The results obtained with immunosuppressive drugs classified by their principal mechanism of action are shown in Table 1, in which the classification of drugs is based on the review by Allison (Allison, 2000).

The MITA demonstrated that dexamethasone (Dex) significantly suppressed IL-2, IL-1β, and IL-8 reporter activities, while cyclosporine A (CyA) and tacrolimus (Tac) suppressed IL-2 and IFN-γ reporter activities but had no effect on IL-1β and IL-8 reporter activities. However, the MITA could not detect the immunosuppressive effects of the alkylating agent cyclophosphamide, of the inhibitors of de novo purine synthesis azathioprine (AZ), mycophenolic acid (MPA) and mizoribine (MZR), and of the inhibitor of pyrimidine and purine synthesis, methotrexate (MT). These data suggest that the MITA correctly evaluates the effects of chemicals on cytokine expression but cannot detect immunotoxicity associated with the inhibition of DNA synthesis and cell division. This drawback has also been reported for other assays, such as the human whole-blood cytokine release assay (HWBCRA) (Langezaal et al., 2002) and the FCP assay (Wagner et al., 2006). On the other hand, the MITA has the advantage that it can discriminate the effects of chemicals on T cells from those on macrophages/dendritic cells.

Table 1. The MITA can detect immunosuppressive effects of representative immunosuppressive drugs

Principal mechanism of action	Drugs	The effects of transcriptional activity			
		IL-2	IFN-γ	IL-1β	IL-8
Immunosuppressing drugs					
Regulation of gene expression	Dexamethasone (Dex)	S	N	S	S
Kinase and phosphatase inhibitors	Cyclosporin A (CyA)	S	S	N	N
	Tacrolimus (Tac)	S	S	N	N
	Rapamycin (RPM)	A	N	N	N
Alkylation	Cyclophosphamide (CP)	N	N	N	N
Inhibition of de novo purine synthesis	Azathioprine (AZ)	N	N	N	N
	Mycophenolic acid (MPA)	A	A	N	N
	Mizoribine (MZR)	N	N	A	A
Inhibition of pyrimidine and purine synthesis	Methotrexate (MTX)	N	A	N	N
Off-label immunosuppressing drugs					
	Sulfasalazine (SASP)	S	S	S	S
	Colchicine	S	N	A	N
	Chloroquine (CQ)	S	N	N	N
	Minocycline (MC)	S	S	N	N
	Nicotinamide (NA)	S	N	S	S
Non-immunomodulatory drugs					
	Acetaminophen (AA)	N	N	N	N
	Digoxin	S	S	N	N
	Warfarin	N	N	S	S

Kimura et al. Toxicol in Vitro 28: 759-769, 2014

*S and A indicates that drugs showed statistically significant suppression in triplicate experiments for each parameter, while N indicates that drugs did not show significant effects.

3-8. The process of validation of the MITA

Although our final goal is to officially validate the MITA for within- and between-laboratory reproducibility and predictivity, in this study, we conducted the validation study for the IL-2 Luc assay as the initial step. Since 2H4 cells used in this validation study is derived from Jurkat cells that contain SLG regulated by the IL-2 promoter, SLO

regulated by the IFN-γ promoter, and SLR regulated by the GAPDH promoter (Saito et al., 2011), this cell line can simultaneously evaluate the effects of chemicals on IL-2 and IFN-γ transcription. However, our previous study demonstrated the significant correlation between the Lowest Observed Effect Levels (LOELs) for the effects of chemicals on the IL-2 luciferase assay and those on the IFN-γ luciferase assay (Kimura et al, 2014). Therefore, we decided to conduct the validation study of only IL-2 Luc assay. Recently, the process of this validation study has been published (Kimura et al., 2020)

3-9. The proposed Adverse Outcome Pathway (AOP) of chemicals that affect IL-2 transcription

Immune dysregulation may have serious impacts on human health, ranging from reduced resistance to infection and neoplasia to allergic and autoimmune conditions. Pivotal immune elements of these diseases are the development of antigen-specific effector T-helper type (Th2) cells, Th1 cells, Th17 cells, and regulatory T cells (Treg cells) that are associated with clinical features and disease progression. Consequently, identifying the immunotoxicity of chemicals requires clarifying their effects on the development of these T cells (reviewed by (Kaiko et al., 2008)).

IL-2 exerts pleiotropic actions on CD4+ T cell differentiation via its modulation of cytokine receptor expression. IL-2 promotes Th1 differentiation by inducing IL-12Rb2 (and IL-12Rb1), promotes Th2 differentiation by inducing IL-4Ra, inhibits Th17 differentiation by inhibiting gp130 (and IL-6Ra), and drives Treg differentiation by inducing IL-2Ra. IL-2 also potently represses IL-7Ra, which decreases survival signals that normally promote cell survival and memory cell development (reviewed by (Liao et al., 2011)). It is therefore conceivable that chemicals that affect IL-2 release by T cells could significantly impact immune function; consequently, we focused on the regulation of IL-2 transcription and attempted to construct an AOP with transcriptional dysregulation of IL-2 as a central key event.

IL-2 mRNA is transcribed after T cell receptor stimulation. Therefore, chemicals that affect any pathway leading to IL-2 transcription after T cell activation can induce dysregulation of IL-2 mRNA and protein expression by T cells. In antigen presentation, T cells are stimulated by T cell receptor (TCR) with co-receptor CD4 or CD8 and CD28. The TCR with CD4 or CD8 recognizes the major histocompatibility complex (MHC)–peptide complex, which results in activation of the SRC kinase Lck and subsequent phosphorylation of immunoglobulin family tyrosine (Y)-based activation motifs (ITAMs) in the CD3 complex (Y-p). This leads to recruitment and phosphorylation of ζ-chain-associated protein (ZAP70), which phosphorylates adaptor proteins, resulting in activation of phospholipase Cγ1 (PLCγ1) and the guanine triphosphatase RAC. PLCγ, in turn, promotes Ca^{2+} mobilization and RAS activation. The combination of these upstream events leads, by complex signaling cascades, to activation of the mitogen-activated protein (MAP) kinases: extracellular signal-regulated kinase (ERK), c-Jun N-terminal kinase (JNK), and p38, as well as phosphatidylinositol 3-kinase (PI3K) and protein kinase B (PKB/Akt). Together, these signals promote different events, including the activation of transcription factors, which result in gene expression and, presumably, T-cell function. On the other hand, CD28 might associate, in its unphosphorylated state, with the serine/threonine phosphatase protein phosphatase 2A (PP2A). Upon T-cell stimulation, CD28 undergoes phosphorylation on its intracellular tyrosine residues (Y), presumably resulting in dissociation from PP2A and recruitment of phosphatidylinositol 3-kinase (PI3K) and growth-factor-receptor-bound protein 2 (GRB2). Activation of PI3K, which induces phosphorylation of phosphatidylinositol (PI) into phosphatidylinositol 3-phosphate (PIP3), might promote activation of protein kinase B (PKB/Akt), followed by activation of nuclear factor-κB (NF-κB), resulting in BCL-XL upregulation that favors T-cell survival. Akt activation might also promote interleukin-2 (IL-2) production. PI3K is negatively regulated by phosphatase and tensin homologue (PTEN). The carboxy-

terminal proline (P)-rich region might promote IL-2 production and proliferation, perhaps by recruiting and activating Lck (reviewed by (Alegre et al., 2001)).

Many chemicals have been reported to affect IL-2 transcription or production. Any component of these signaling cascades can be a potential target of these chemicals, but the mechanism by which they affect IL-2 transcription or production remains largely unknown.

Based on recent advances in immunology, we tentatively propose the following AOP for immunosuppression focusing on IL-2 transcription. Figure 7 shows the AOP with representative chemicals that affect IL-2 transcription. From 2001 to 2017, 54 chemicals were reported to augment IL-2 gene or protein expression in human and 60 chemicals had this effect in mice, while 65 chemicals in human and 47 chemicals in mice were reported to decrease IL-2 gene or protein expression, as determined by a PubMed search.

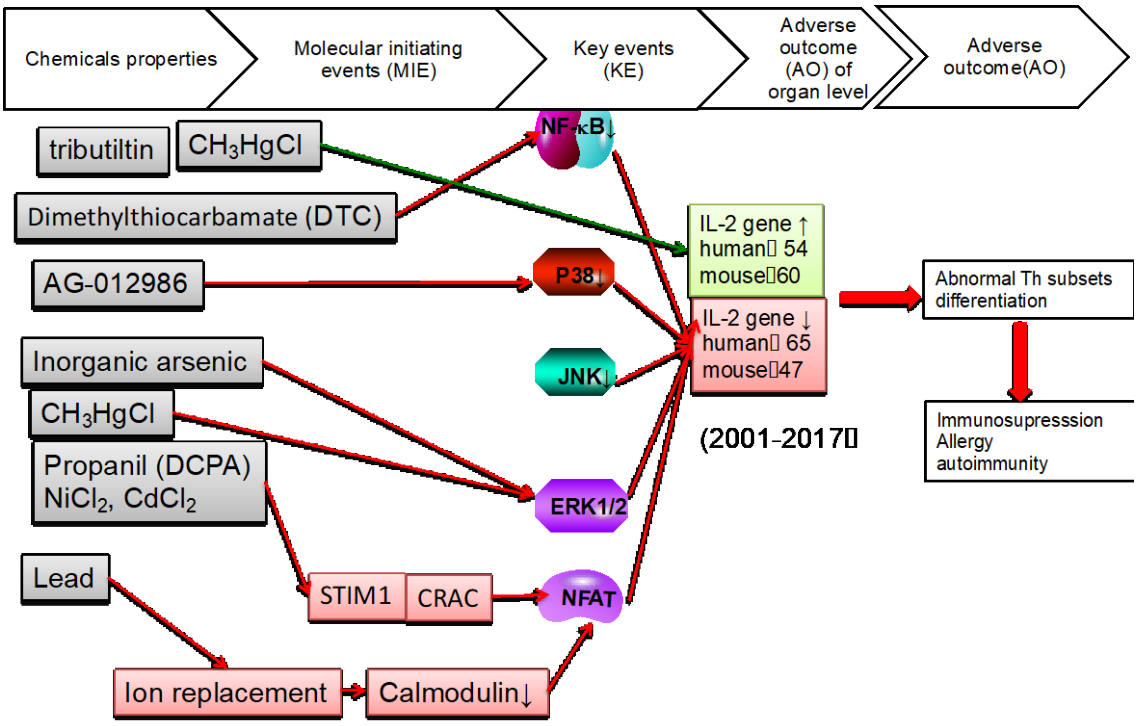


Fig. 7. The proposed AOP for dysregulation of Th subset differentiation triggered by disrupted IL-2 transcription.

4. Test method and modification

4-1. IL-2 reporter cell, 2H4

The Jurkat human acute T lymphoblastic leukemia cell line kindly provided by Professor Kazuo Sugamura, Department of Microbiology, Tohoku University School of Medicine, was cultured in RPMI-1640 (Sigma-Aldrich, St. Louis, MO) containing Antibiotic-Antimycotic (Invitrogen) and 10% Hyclone^{TM} fetal calf serum (Thermo Fisher Scientific, Waltham, MA) (Jurkat growth medium) at 37°C with 5% CO_2. The luciferase reporter assay system was constructed using 3 luciferases that emit green light (Stable luciferase green; SLG), orange light (Stable luciferase orange; SLO), and red light (Stable luciferase red; SLR) using a single substrate. Namely, we constructed three luciferase vectors, pSLG-test/Hyg^r, pSLO-test/Neo^r, and pSLR-test/Pur^r, by ligating the *Bam*HI/*Sac*I site of resistant gene vectors containing one of three resistant genes, hygromycin (SLG), neomycin (SLO) or puromycin (SLR), SV40 promoter, and HSVtk polyA into luciferase gene vectors, pSLG-test, pSLO-test and pSLR-test (Toyobo, Osaka, Japan), respectively. The activities of the luciferases can be measured simultaneously and quantitatively with optical filters. This system can rapidly and easily monitor the expression of multiple genes (Nakajima et al., 2005; Noguchi et al., 2008).

4-2. Chemical treatment of 2H4 cells and measurement of luciferase activity

Based on previous reports(Saito et al., 2011; Takahashi et al., 2011), 2H4 cells (2×10^5 cells/50 µl/well) in 96-well black plates (Greiner Bio-One GmbH, Frickenhausen, Germany) were pretreated with different concentrations of individual chemicals for 1 h. The 2H4 cells were then stimulated with 25 nM PMA and 1 µM ionomycin (PMA/Io) for 6 h. Three luciferase activities (SLG luciferase activity (SLG-LA), SLO luciferase activity (SLO-LA), and SLR luciferase activity (SLR-LA)) were simultaneously determined using a microplate-type luminometer with a multi-color detection system (Phelios; Atto Co., Tokyo, Japan) and Tripluc luciferase assay reagent (TOYOBO Co., Ltd., Osaka, Japan) according to the manufacturers' instructions. Use of

the 2H4 cell line enabled measurement of SLO-LA driven by the IL-2 promoter (IL2LA), SLG-LA driven by the INF-γ promoter (IFNLA), and SLR-LA driven by GAPDH (GAPLA) in 2H4 cells. In this validation study, however, we just used the IL2LA and GAPLA and ignored IFNLA because there was a significant correlation between LOELs for the effects on the IL2LA and those on the IFNLA (Kimura et al., 2018). We accounted for the variation in cell number and cell viability after chemical treatment by normalizing the data for IL2LA (nIL2LA) by dividing IL2LA with GAPLA in the 2H4 cells. In addition, we calculated % suppression, % augmentation, and Inh-GAPLA as follows:

% suppression = (nIL2LA of 2H4 cells treated with chemicals/nIL2LA of non-treated 2H4 cells) x 100;

% augmentation = (1-(nIL2LA of 2H4 cells treated with chemicals/nIL2LA of non-treated 2H4 cells)) x 100;

Inh-GAPLA = GAPLA of 2H4 cells treated with chemicals/GAPLA of untreated cells.

Definitions of these terms are provided in Table 2.

Table 2. Definition of the parameters in the IL-2 Luc assay.

Abbreviations	Definition
IL-2 Luc assay	IL-2 luciferase assay
GAPLA	SLR luciferase activity reflecting GAPDH promoter activity
IL2LA	SLO luciferase activity reflecting IL-2 promoter activity of 2H4 cells
IFNLA	SLG luciferase activity reflecting IFN-γ promoter activity of 2H4 cells
nIL2LA	IL2LA/GALA of 2H4 cells
nIFNLA	IFNLA/GALA of 2H4 cells
% suppression	(nIL2LA of 2H4 cells treated with chemicals/ nIL2LA of non-treated 2H4 cells) x 100
% augmentation	(1-(nIL2LA of 2H4 cells treated with chemicals/ nIL2LA of non-treated 2H4 cells)) x 100
CV05	The lowest concentration of the chemical at which Inh-GAPLA becomes < 0.05.
Inh-GAPLA	GAPLA of 2H4 cells treated with chemicals /GAPLA of untreated cells.

4-3. Criteria to determine the effects of chemicals on T cells

During the validation study, we modified the criteria to determine the effects of chemicals on T cells to determine the criteria for the MITA.

We used the following Criteria 1 in our first publication describing the MITA. Three independent experiments were conducted for each chemical. For each experiment, a one-way ANOVA test followed by Dunnett's post hoc test was used to evaluate statistical significance. If chemicals showed statistically significant immunosuppression or immunostimulation in 3 experiments, they were judged as immunosuppressive or immunostimulatory drugs, respectively. If chemicals showed statistically significant immunosuppression or immunostimulation in only 2

independent experiments, they were judged as potential immunosuppressive or immunostimulatory drugs, respectively. If not, they were judged as ineffective. Then, for potential immunosuppressive or immunostimulatory drugs, we selected their percent suppression or percent augmentation (negative percent suppression) in 3 experiments that showed the most significant change, calculated their percent suppression or percent augmentation, and statistically compared suppression or augmentation by the chemicals with that of the vehicle control in 3 different experiments by the Student's t-test. Only when chemicals demonstrated statistical significance were they judged as immunosuppressive or immunostimulatory, respectively(Kimura et al., 2014).

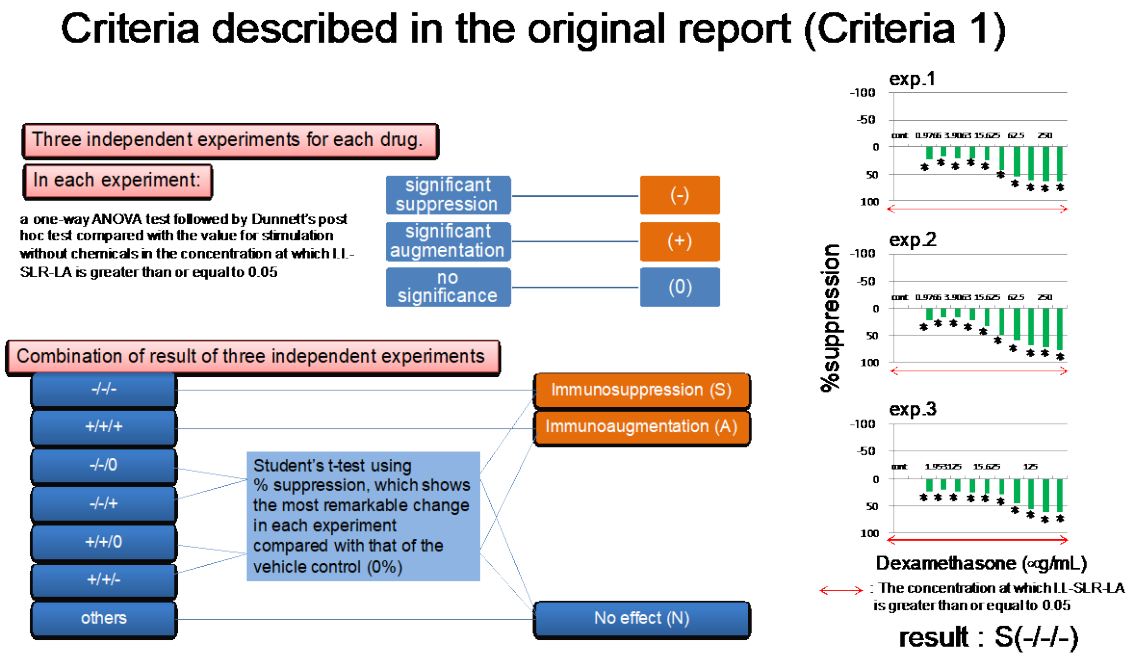


Fig.8　Criteria 1 in the original report

After the pre-validation study, in addition to the original criteria (Criteria 1, Fig.8), two new criteria were proposed by the statistician (Criteria 2, Criteria 3). These 3 criteria were used temporarily and one of these criteria would be adopted after the Phase I validation study.

4-4. Bioluminescence system

In a typical dual-reporter assay, firefly luciferase from *Photinus pyralis* (FLuc) is used as the experimental reporter and Renilla luciferase is used as the internal control reporter. This internal control reporter connects to a constitutively expressed promoter, such as the herpes simplex virus thymidine kinase promoter, cytomegalovirus (CMV) immediate-early promoter, or simian virus 40 (SV40) promoter. This assay system is commercialized as a Dual-Luciferase Reporter Assay System by Promega Corporation. In this system, both luciferase activities are measured sequentially from single extracts on the basis of their bioluminescent substrate specificity. Firefly luciferase activity is measured first by adding firefly D-luciferin, and then Renilla luciferase activity is measured by adding coelenterazine (another name for Renilla luciferin), with concomitant quenching of firefly luciferase luminescence. Finally, firefly luciferase activity is normalized by Renilla luciferase activity as the promoter activity (Michelini et al., 2014; Nakajima and Ohmiya, 2010; Roda et al., 2004).

An alternative chemical test using a cell-based assay requires the analysis of a large number of samples. It is therefore preferable to use an improved assay system whereby gene expression can be monitored simultaneously in a one-step reaction in single extracts. Beetle luciferases emit red luminescence during reaction, compared to the green emitted by firefly D-luciferin. The two colors can be divided using an optical filter. The dual color-reporter assay is based on the color difference between beetle and firefly luciferases and is sold commercially as the Tripluc Reporter Assay System by TOYOBO (Nakajima et al., 2004; Nakajima et al., 2005).

In the IL-2 Luc assay, the multicolor luciferase assay system (Nakajima et al. 2005) consisted of a green-emitting luciferase (SLG; lmax = 550 nm) for the gene expression

of the IL-2 promoter, an orange-emitting luciferase (SLO; lmax = 580 nm) for the gene expression of the IFN-γ promoter, and a red-emitting luciferase (SLR; lmax = 630 nm) for the gene expression of the internal control promoter, GAPDH.

The three luciferases emit different colors upon reacting with firefly D-luciferin and their luminescence is measured simultaneously in a one-step reaction by dividing the emission from the assay mixture using an optical filter (Nakajima et al., 2005). First, the total relative light units (F0) are measured in the absence of the filters. Then, the F1 and F2 values that passed through the R56 filter (>560-nm long-pass filters) or the R60 filter (>600-nm long-pass filters), respectively, is measured. The three luciferase activities are calculated using the simultaneous equation shown below by substituting the F0, F1 and F2 values. In this equation, G, O and R are the activities of the green-, orange- and red-emitting luciferases, respectively, κG_{R56}, κO_{R56} and κR_{R56} are the transmission coefficients of the green-, orange- and red-emitting luciferases of the R56 filter, respectively, κG_{R60}, κO_{R60} and κR_{R60} are the transmission coefficients of the green-, orange- and red-emitting luciferases of the R60 filter, respectively.

$$\begin{pmatrix} F0 \\ F1 \\ F2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ \kappa G_{R56} & \kappa O_{R56} & \kappa R_{R56} \\ \kappa G_{R60} & \kappa O_{R60} & \kappa R_{R60} \end{pmatrix} \begin{pmatrix} G \\ O \\ R \end{pmatrix}$$

Luminescence activity is measured using a filtered 96-well microplate luminometer (for example, Phelios (ATTO, Tokyo, Japan), Tristan 941 (Berthold, Bad Wildbad, Germany), and the ARVO series (PerkinElmer, Waltham, MA). It is necessary to calibrate the luminometer in each experiment to ensure reproducibility (Niwa et al., 2010). Recombinant green-, orange- and red-emitting luciferases are available for this calibration.

5. Validation Management Structure

5-1. Validation Management Team (VMT)

Trial Coordinator:	Hajime Kojima (Japanese Center for the Validation of Alternative Methods (JaCVAM), National Institute of Health Sciences (NIHS), Kawasaki, Japan), VMT trial coordinator, Chemical supplier and Management of quality control
Lead laboratory:	Setsuya Aiba (Tohoku University, Miyagi, Japan), Developer of this assay, Test method, expertise underlying science Yutaka Kimura (Tohoku University, Miyagi, Japan)
International expert members	
EU liaison:	Emanuela Corsini (Milan Univ., Italy), Test system expertise, validation expertise, immunotoxicity expertise Erwin L. Roggen (3Rs Management and Consulting ApS, Denmark), Test system expertise, validation expertise, immunotoxicity expertise
ICCVAM liaison:	Dori Germolec (NTP/NIEHS, USA), Immunotoxicity expertise
JSIT liaison:	Tomoaki Inoue (Chugai Pharmaceutical Co., Ltd.), Immunotoxicity expertise
Data management team:	Takashi Omori (Kobe University, Kobe, Japan), Data analysis, biostatistics dossier
Chemical Selection Committee	Setsuya Aiba (Tohoku University) Yutaka Kimura (Tohoku University Hajime Kojima (JaCVAM) Emanuela Corsini (Milan Univ)

	Erwin L. Roggen (3Rs Management and Consulting ApS) Dori Germolec (NTP/NIEHS) Tomoaki Inoue (Chugai Pharmaceutical Co., Ltd.)
Participating Test Facilities	Test Facility 1: Hatano Res. Inst., FDSC, Study Director (SD): Kohji Yamakage Test Facility 2: AIST, Tsukuba, SD: Rie Yasuno Test Facility 3: AIST, Takamatsu, SD: Yoshihiro Nakajima

5-2. Management office

Hajime Kojima (JaCVAM)
3-25-26 Yodomimati Kawasaki, Kawasaki, 210-9501
TEL: +81-44-270-6600
h-kojima@nihs.go.jp

5-3. Meetings

27-28/1/2016 (Mitoya, Sendai, Japan)
1st International VMT Meeting
Subjects: Kick-off meeting for the MITA assay
VMT members: Corsini, E., Roggen, E., Germolec, D.(telephone), Inoue, T., Kageyama, S., Aiba, S., Kimura, Y., Yamakage, K., Watanabe, M., Kobayashi, M., Yasuno, R., Ohmiya, Y., Omori, T., Kojima, H., Tanabe, S., Venti, S.
Participating laboratories: AIST(Tsukuba), HRI

13/9/2016 (Skype-meeting)
Meeting by Skype
Subjects: Result of the phase 0 study and proposal of the revised protocol

VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T.,
Aiba, S., Kimura, Y., Omori, T., Kojima, H.

4-5/2/2017 (Nayamachi community hall, Kyoto, Japan)
2nd International VMT Meeting
Subjects: Validation results, discussion and suggestion
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T.,
Aiba, S., Kimura, Y., Yamakage, K., Watanabe, M., Kobayashi, M.,
Yasuno, R., Nakajima, Y., Omori, T., Mori, A., Kobayashi, M.,
Kojima, H., Venti, S.
Participating laboratories: AIST(Tsukuba), HRI, AIST(Takamatsu)

18-19/11/2017 (Umeda Center Building, Osaka, Japan)
3rd International VMT Meeting
Subjects: Validation results, discussion and suggestion
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T.,
Aiba, S., Kimura, Y., Yamakage, K., Watanabe, M., Kobayashi, M.,
Yasuno, R., Nakajima, Y., Omori, T., Mori, A., Kobayashi, M.,
Kojima, H., Venti, S.
Participating laboratories: AIST(Tsukuba), HRI, AIST(Takamatsu)

29/3/2018 (Skype-meeting)
Meeting by Skype
Subjects: Proposal of the revised protocol
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T.,
Aiba, S., Kimura, Y., Omori, T., Kojima, H.

10/4/2018 (telephone-meeting)

Meeting by telephone

Subjects: Understanding the unexpected results in the IL-2 Luc assay

VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T., Aiba, S., Kimura, Y., Omori, T., Kojima, H.

4-6/10/2018 (Kobe Univ., Kobe, Japan)

4th meeting for the MITA Validation study

Subjects: Validation report for the IL-2 assay

VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T. Aiba, S., Kimura, Y., Yamakage, K., Watanabe, M., Yasuno, R., Nakajima, Y., Omori, T., Takagi, Y., Mashimo, N., Kado, Y., Kojima, H., Venti, S.

Participating laboratories: AIST(Tsukuba), HRI, AIST(Takamatsu)

6. Study Design (Appendix 12)

The aim of this phase is to (pre)validate the IL-2 Luc assay method to assess transferability and inter-laboratory variability so that this test can be used to screen for immunotoxic chemicals.

The validation study (Phase I and Phase II trials) was conducted by 3 laboratories, based on the study design and schedule shown in Tables 3 and 4 and using the test chemicals shown in Tables 5 and 6. The methods were described above in section 4: ’Test Method 4.1 IL-2 Luc assay’, and the precise protocol is described below in section 8: ‘Protocol 8.2 Protocol for the IL-2 Luc assay’ in Tables 7-9.

Table 3. The number of chemicals analyzed in the validation study

Studies	Within-Laboratory	Between-laboratories	Predictivity
I	5	5	5
II		20	20
Total	5	25	25

7. Test Chemicals

The selection process for the test chemicals for the IL-2 Luc assay validation study is described below.

In addition, the chemical categories or physical state and chemical properties (e.g., solid, liquid, etc.) are included in the tables of these test chemicals in order to investigate the applicable domain.

Table 4. Breakdown of the IL-2 Luc assay validation study

Phase	The number of the test substances	The number of the repetitions	Examination	Date of experiment start
Pre	5	1	Between- laboratory transferability (Non-coded)	July, 2016
I	5	3	Within- and between- laboratory reproducibility (Coded)	September, 2016
II	20	1	Between- laboratory reproducibility and predictivity (Coded)	May, 2017

7-1. Basic rule for chemical selection

The selection of test chemicals by the Chemical Selection Committee (CSC) in the VMT was based on published papers on *in vivo* immunotoxicity tests and validation studies for *in vitro* alternative assays on immunotoxicity test methods.

7-1-1. The applied selection criteria

- information on mode/site of action
- coverage of a range of relevant chemical classes and product classes
- quality and quantity of reference data (*in vivo* and *in vitro*)
- high-quality data derived from animal and (if available) human studies
- information on interspecies variations (for example: variability with regard to the uptake of chemicals, metabolism, etc.)
- coverage of a range of toxic effects/potencies
- chemicals that do not require metabolic activation
- appropriate negative and positive controls

- physical and chemical properties (feasibility of use in the experimental set-up as implicated by the CAS No.)
- single chemical entities or formulations of known high purity
- availability
- cost

In the first phase of the selection procedure, the CSC identified and collected several existing lists of potential chemical immunotoxicants, such as NTP IMMUNOTOX, EPA candidate list. An extensive literature search was performed by the CSC in order to ensure that all the pre-selected chemicals fulfilled the selection criteria described above. In addition, it was decided that at least 20% of the total chemicals to be tested should provide negative results (i.e., not immunotoxic) in order to increase the statistical power of the data analysis.

7-1-2. Chemical Acquisition, Coding and Distribution

Laboratory transferability, and within- and between-laboratory reproducibility and predictivity, in all test facilities were assessed using coded chemicals. Coding was supervised by JaCVAM, in collaboration with CSC. CSC was responsible for coding and distributing the test chemicals, references, and controls for the validation study.

7-1-3. 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 hazards identification and exposure controls/personal protection for each chemical. 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.

7-2. Pre-validation study

Transferability of this assay was checked using five non-coded chemicals (2-aminoanthracene, citral, chloroquine diphosphate salt, dexamethasone and methylmercury(II) chloride) (Appendix 1) in 4 test facilities, including the lead laboratory. These chemicals were selected by the CSC.

7-3. Validation study -Phase I trial

Within- and between-laboratory reproducibility of this assay was checked using 5 coded chemicals in 3 test facilities. These chemicals were selected by CSC based on the in-house data set of the lead laboratory. The chemicals were coded by JaCVAM as shown in Table 5 (Appendix 2) and distributed to the test facilities.

Table 5. Chemical code list on the phase I validation trial for IL-2 Luc assay

No.	Chemical	CASRN	MW	Supplier	Catalog No.	Content	Physical characteristics	Lot	Storage	Purity	LabA	LabB	LabC	LabD
											TOHOKU univ.	AIST-TSUKUBA	FDSC	AIST-SHIKOKU
1	Dibutyl phthalate	84-74-2	278.34	Wako	021-06936	500mL	Liquid	TLN0112	RT	98.0+% (Capillary GC)	MIA003A	MIB014A	MIC027A	MID036A
											MIA004B	MIB017B	MIC026B	MID033B
											MIA007C	MIB016C	MIC023C	MID034C
2	Hydrocortisone (for Cell Culture)	50-23-7	362.46	Wako	080-10194	50g	Solid	SAH3714	RT	97% (HPLC)	MIA005A	MIB017A	MIC029A	MID038A
											MIA007B	MIB019B	MIC028B	MID035B
											MIA009C	MIB018C	MIC025C	MID037C
3	Lead(II) acetate trihydrate (Deleterious substances)	6080-56-4	379.33	Sigma-Aldrich	316512-100G	100g	Solid	09901TS	RT	99.999% trace metals basis	MIA007A	MIB018A	MIC021A	MID310A
											MIA008B	MIB011B	MIC210B	MID037B
											MIA001C	MIB110C	MIC027C	MID038C
4	Zinc dimethyldithiocarbamate (DMDTC)	137-30-4	305.82	Kanto Chemical	48028-31	25g x2	Solid	403N2204	RT	>95.0% (T)	MIA009A	MIB110A	MIC023A	MID037A
											MIA010B	MIB013B	MIC027B	MID039B
											MIA003C	MIB017C	MIC029C	MID310C
5	Nickel (II) sulfate hexahydrate	10101-97-0	262.85	Wako	146-01171	100g	Solid	LKQ2263	RT	99.0-102.0% (as NiSO4・6H2O) (Titration)	MIA001A	MIB012A	MIC025A	MID034A
											MIA002B	MIB015B	MIC024B	MID031B
											MIA005C	MIB014C	MIC021C	MID032C

7-4. Validation study -Phase II trial

Between-laboratory reproducibility of this assay was checked using 20 coded chemicals in 3 test facilities. The chemicals were coded by JaCVAM as shown in Table 6 (Appendix 3) and distributed to the test facilities.

Table 6. Chemical code list on the phase II validation trial for IL-2 Luc assay

	Chemical	Cas.no.	LabA	LabB	LabC	LabD	Note	State	Storage	Supplier	Lot
			TOHOKU univ.	AIST-TSUKUBA	FDSC	AIST-SHIKOKU					
1	2,4-Diaminotoluene	95-80-7	MIA401	MIB515	MIC618	MID702	Deleterious	S	RT	Wako	CDF0347
2	Benzo(a)pyrene	50-32-8	MIA413	MIB516	MIC601	MID703		S	RT	TCI	M8DFD
3	Cadmium chloride	10108-64-2	MIA403	MIB502	MIC602	MID714	Deleterious	S	RT	Wako	DEE3332
4	Dibromoacetic acid	631-64-1	MIA406	MIB518	MIC610	MID720		S	RT	ALDRICH	BCBR5175V
5	Diethylstilbestol	56-53-1	MIA420	MIB509	MIC611	MID711		S	RT	SIGMA	BCBR9766V
6	Diphenylhydantoin	630-93-3	MIA412	MIB510	MIC615	MID704		S	RT	SIGMA	SLBB3874
7	Ethylene dibromide	106-93-4	MIA407	MIB507	MIC605	MID705	Deleterious	L	RT	Wako	KWG5479
8	Glycidol	556-52-5	MIA408	MIB505	MIC607	MID712		L	2-8℃	ALDRICH	MKBX5752V
9	Indomethacin	53-86-1	MIA409	MIB508	MIC609	MID715		S	RT	SIGMA	122K0718
10	Isonicotinic Acid Hydrazide (Isoniazid)	54-85-3	MIA411	MIB517	MIC612	MID707		S	RT	Fluka	SLBF8371V
11	Nitrobenzene	98-95-3	MIA402	MIB519	MIC603	MID701	Deleterious	L	RT	Sigma-Aldrich	SHBG5577V
12	Urethane, Ethyl carbamate	51-79-6	MIA415	MIB520	MIC604	MID719		S	RT	Sigma-Aldrich	WXBC3505V
13	Tributyltin chloride	1461-22-9	MIA404	MIB506	MIC613	MID713	Deleterious	L	RT	TCI	2442A-IQ
14	Perfluorooctanoic acid	335-67-1	MIA414	MIB514	MIC614	MID718		S	RT	TCI	O3U70
15	Dichloracetic acid	79-43-6	MIA416	MIB511	MIC606	MID716	Deleterious	L	RT	Sigma-Aldrich	SHBH3492V
16	Toluene	108-88-3	MIA417	MIB512	MIC616	MID706	Deleterious	L	RT	Sigma-Aldrich	J5136
17	Acetonitril	75-05-8	MIA405	MIB501	MIC617	MID708	Deleterious	L	RT	Wako	KWH4805
18	Mannitol	69-65-8	MIA418	MIB503	MIC619	MID717		S	RT	Wako	LKP4362
19	Vanadium pentoxide	1314-62-1	MIA419	MIB504	MIC608	MID709	Deleterious	S	RT	Wako	SAE6958
20	o-Benzyl-p-chorolophenol	120-32-1	MIA410	MIB513	MIC620	MID710		S	RT	Wako	KPQ0988

7-5. Acceptance criteria

The within-laboratory reproducibility for the all test facilities was done by an independent biostatistical analysis using coded five chemicals, under the VMT. The proportion of concordance should be equal or more than 80% as tentative acceptance criteria for phase I study.

Twenty-five coded test items have been selected to confirm the between-laboratory reproducibility in the phase I and II study. At the end of the testing, the test facilities will submit a QC certified copy of whole study dossier to the trial coordinator (study plan in GLP principle, raw data, records and data analysis, study report in GLP principle). The proportion of concordance between-laboratory reproducibility should be equal or more than 80% as acceptance criteria.

8. Protocols

Overview of the IL-2 Luc assay

An overview of the IL-2 Luc assay is shown in Fig. 9. In addition, the final protocol of the present test (version 011.1E) is provided as attached Appendixes 4 and 5, and the procedures are described in detail below.

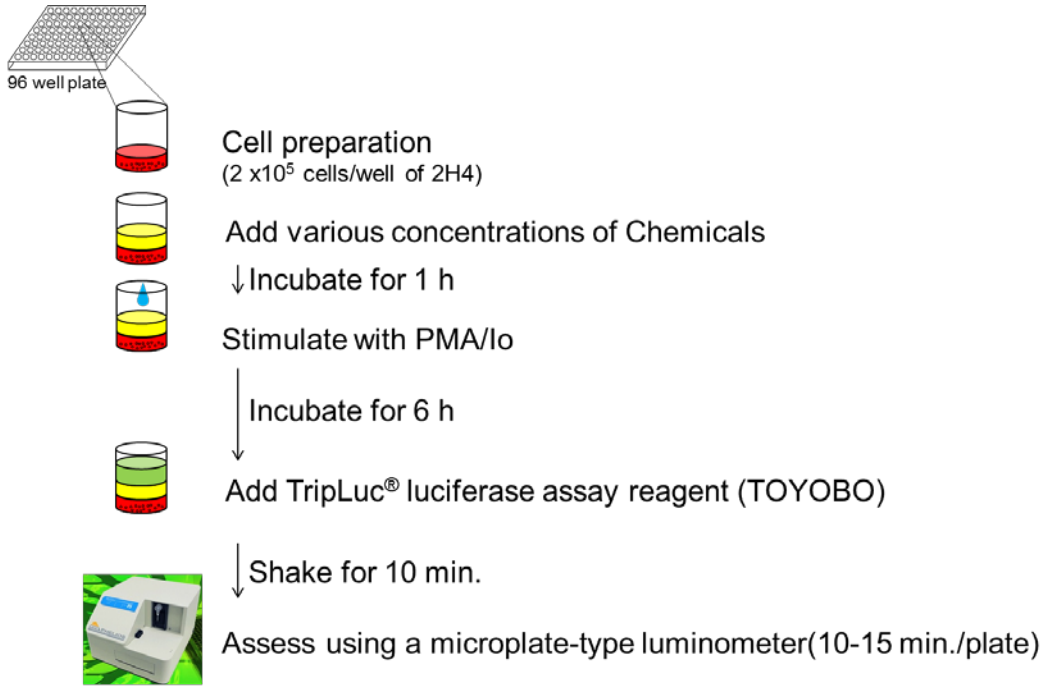


Fig. 9. Overview of the IL-2 Luc assay

8-1 Cells

・ 2H4 (IL2-SLG、IFNγ-SLO、GAPDH-SLR)

The human acute T lymphoblastic leukemia cell line Jurkat was obtained from the ATCC. A Jurkat-derived IL-2 and IFN-γ reporter cell line, 2H4, that harbors the SLG, SLO and SLR luciferase genes under the control of the IL-2, IFN-γ and GAPDH promoters, respectively, was established by Tsuruga Institute of Biotechnology, TOYOBO Co. Ltd., Fukui, Japan. (Saito et al. 2011)

8-2. Protocol for the IL-2 Luc assay

8-2-1. Reagents and equipment (Appendix 6)

The following reagents and equipment were used.

For maintenance of 2H4 cells

・ RPMI-1640 (GIBCO Cat#11875-093, 500 mL)

・ FBS (Biological Industries Cat#04-001-1E Lot: 715004)

- Antibiotic-Antimycotic (GIBCO Cat#15240-062)
- HygromycinB (CAS:31282-04-9, Invitrogen Cat#10687-010)
- G418 (CAS:108321-42-2, Nacalai Tesque Cat#16513-84)
- Puromycin (CAS:58-58-2, InvivoGen Cat#ant-pr-1)

For chemical exposure, stimulation and solvents

- Ionomycin (CAS:56092-82-1, Sigma Cat#I0634)
- Phorbol 12-myristate 13-acetate (PMA) (CAS:16561-29-8, Sigma Cat#P8139)
- Ethanol (e.g., Wako Cat#057-00456)
- Dimethyl sulfoxide (DMSO) (CAS:67-68-5, Sigma Cat#D5879)
- Distilled water (GIBCO Cat#10977-015)

For measurement of luciferase activity

- Triplu[®] Luciferase assay reagent (TOYOBO Cat#MRA-301)

8-2-2. Culture medium

Various culture media were used depending on the purpose of the cell culture.

Table 7. A medium: for maintenance of 2H4 cells (500 mL, stored at 2-8°C)

Reagent	Company	Concentration	Final concentration in medium	Required amount
RPMI-1640	GIBCO #11875-093	-	-	440 mL
FBS	Biological Industries Cat#04-001-1E Lot: 715004	-	10 %	50 mL
Antibiotic-Antimycotic	GIBCO #15240-062	100×	1×	5 mL
Puromycin	InvivoGen #ant-pr-1	10 mg/mL	0.15 μg/mL	7.5 μL
G418	Nacalai Tesque #16513-84	50 mg/mL	300 μg/mL	3 mL
HygromycinB	Invitrogen #10687-010	50 mg/mL	200 μg/mL	2 mL

Table 8. B medium: for luciferase assay (30 mL, stored at 2-8°C)

Reagent	Company	Concentration	Final concentration in medium	Required amount
RPMI-1640	GIBCO #11875-093	-	-	27 mL
FBS	Biological Industries Cat#04-001-1E Lot: 715004	-	10 %	3 mL

Table 9. C medium: for thawing 2H4 cells (30 mL, stored at 2-8°C)

Reagent	Company	Concentration	Final concentration in medium	Required amount
RPMI-1640	GIBCO #11875-093	-	-	26.7 mL
FBS	Biological Industries Cat#04-001-1E Lot: 715004	-	10 %	3 mL
Antibiotic-Antimycotic	GIBCO #15240-062	100×	1×	0.3 mL

8-2-3. Cell line

The Jurkat human acute T lymphoblastic leukemia cell line (ATCC, Manassas, VA, USA), was cultured in RPMI-1640 (Sigma-Aldrich, St. Louis, MO, USA) with Antibiotic-Antimycotic (Invitrogen, Carlsbad, CA, USA) and 10% Hyclone™ fetal calf serum (Thermo Fisher Scientific, Wilmington, NC, USA) (Jurkat growth medium) at 37 °C with 5% CO_2. The luciferase reporter assay system was constructed using three different luciferases, SLG, SLO and SLR, that emit green, orange, and red light, respectively, with a single substrate. In brief, we constructed three luciferase vectors, pSLG-test/Hygr, pSLO-test/Neor, and pSLR-test/Purr, by ligating the BamHI/SacI site of resistant gene vectors containing one of the three resistant genes, hygromycin (SLG), neomycin (SLO) or puromycin (SLR), the SV40 promoter, and HSVtk polyA into the luciferase gene vectors, pSLG-test, pSLO-test and pSLRtest (Toyobo, Osaka, Japan), respectively. The activities of the luciferases can be measured simultaneously and quantitatively using optical filters. This system can rapidly and easily monitor multiple gene expression (Nakajima et al., 2005; Noguchi et al., 2008). Promoter cloning was

carried out as follows. The IL-2 promoter construct containing nt －3006 to +286, the IFN-γ promoter construct containing nt －4971 to +111, and the GAPDH promoter construct containing nt －1373 to +128 from transcription initiation sites that were identified using DBTSS (http://dbtss.hgc.jp/), were amplified from genomic DNA by PCR using KOD-Plus- ver. 2 (Toyobo) for the IL-2 promoter or KOD-Plus- (Toyobo) for the IFN-γ and GAPDH promoters and specific primers. The IL-2 promoter, IFN-γ promoter, or GAPDH promoter was ligated into pSLG-test/Hygr, pSLOtest/Neor or pSLR-test/Purr vectors that had been digested with MluI and XhoI, MluI and SalI, or MluI and EcoRI, respectively. Before transfection, we confirmed the sequence of the 5' and 3' regions of each promoter using a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). IL-2, IFN-γ and GAPDH reporter plasmids (1 μg) were transfected into Jurkat T cells ($5x10^5$ cells) using SuperFect (Qiagen, Valencia, CA, USA). After transfection, cells were cultured in Jurkat growth medium containing 200 μg/ml hygromycin (Invitrogen), 300 μg/ml G418 (Nacalai tesque, Kyoto, Japan) and 0.15 μg/ml puromycin (InvivoGen, San Diego, CA, USA) for selection. After repeated limiting dilution, we established a stable cell line (2H4 cells) in Fig.10.

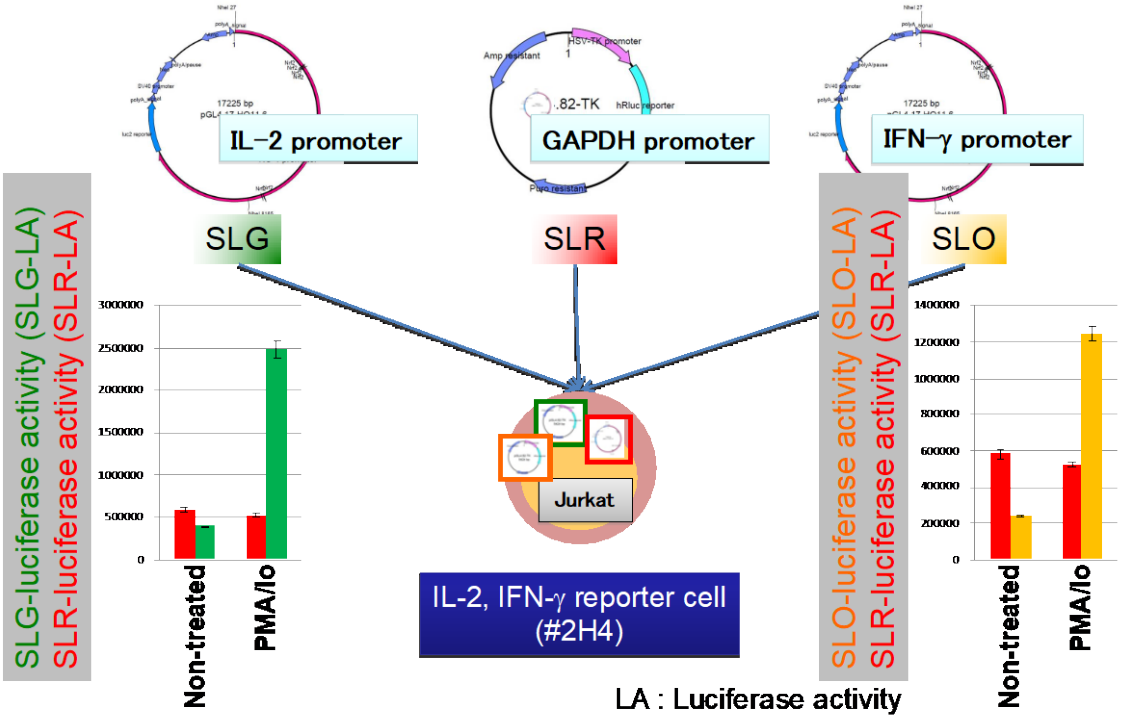


Fig. 10. IL-2 reporter cell, 2H4

8-2-4. Thawing of 2H4 cells

Pre-warm 9 mL of C medium in a 15 mL polypropylene conical tube in a 37°C water bath (for centrifugation) and 15 mL of C medium in a T-75 flask at 37°C in a 5% CO_2 incubator (for culture).

Thaw frozen cells ($2x10^6$ cells/0.5 mL of freezing medium) in a 37°C water bath, then add to a 15 mL polypropylene conical tube containing 9 mL of pre-warmed C medium. Centrifuge the tube at 120-350 x *g* at room temperature for 5 min, discard the supernatant, and resuspend in 15 mL of pre-warmed C medium in a T-75 flask. Cells are incubated at 37°C, 5% CO_2.

8-2-5. Maintenance of 2H4 cells

Pre-warm the A medium in a T-75 Flask at 37°C in a 5% CO_2 incubator. The culture medium should be changed to the pre-warmed A medium 3 or 4 days after thawing. At that time, count the number of cells, centrifuge the tube at 120-350 x *g* at room temperature for 5 min, discard the supernatant, and resuspend in pre-warmed the A

medium in a T-75 Flask. Cells are passaged at $3x10^5$/mL and incubated at 37°C, 5% CO_2.

The interval between subcultures should be 3~4 days. Cells can be used between one and six weeks after thawing.

The lead laboratory has examined how long 2H4 cells could be cultured without losing their reactivity to PMA/Io. 2H4 cells maintained their response to PMA/Io up to 16 weeks or 35 passages.

8-2-6. Preparation of cells for assay

A cell passage should be done 2-4 days before the assay.

Use cells between 1 and 6 weeks after thawing.

Pre-warm the B medium in a 37°C water bath. Count the number of cells and collect the number of cells needed (2.0×10^7 cells for two chemicals are required, but to have some leeway, 3.0×10^7 cells for two chemicals should be prepared), centrifuge the tube at 120-350 x *g*, 5 min. Resuspend in pre-warmed the B medium at a cell density of 4×10^6/mL. Transfer the cell suspension to a reservoir (Thermo Scientific), and add 50 µL of cell suspension to each well of a 96 well µclear black plate (flat bottom) using an 8 channel or 12 channel pipetman (Gison, Inc, Middleton, WI, USA). (cf. Figure 11)

flat-bottom black	1	2	3	4	5	6	7	8	9	10	11	12
A	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
B	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
C	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
D	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
E	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
F	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
G	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL
H	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL	2H4 2x10^5 B medium 50uL

Fig. 11. Components in each well of 96-well plates after cell preparation.

8-2-7. Preparation of chemicals and cell treatment with chemicals

In Fig. 12, water soluble chemicals were dissolved in distilled water at a concentration of 25 mg/mL. If the chemicals were soluble at 25 mg/mL, then 50 mg/mL solutions were prepared if their solubility was sufficient. If they were not soluble at 50 mg/mL, then 25 mg/mL was judged the highest soluble concentration. If the chemicals were soluble at 50 mg/mL, then 100 mg/mL solutions were prepared if their solubility was sufficient. If they were not soluble at 100 mg/mL, then 50 mg/mL was judged the highest soluble concentration. If they were soluble at 100 mg/mL, then 100 mg/mL was judged the highest soluble concentration.

Chemicals not soluble in water were dissolved in DMSO at 500 mg/mL. If they were not soluble at 500 mg/mL, the highest soluble concentration was determined by diluting the suspension from 500 mg/mL by a factor of 2 with DMSO. Sonication and vortex mixing were used if needed and the attempt to dissolve the chemical continued for

at least 5 minutes. All dissolved chemicals were used within 4 hours of being dissolved in distilled water or DMSO.

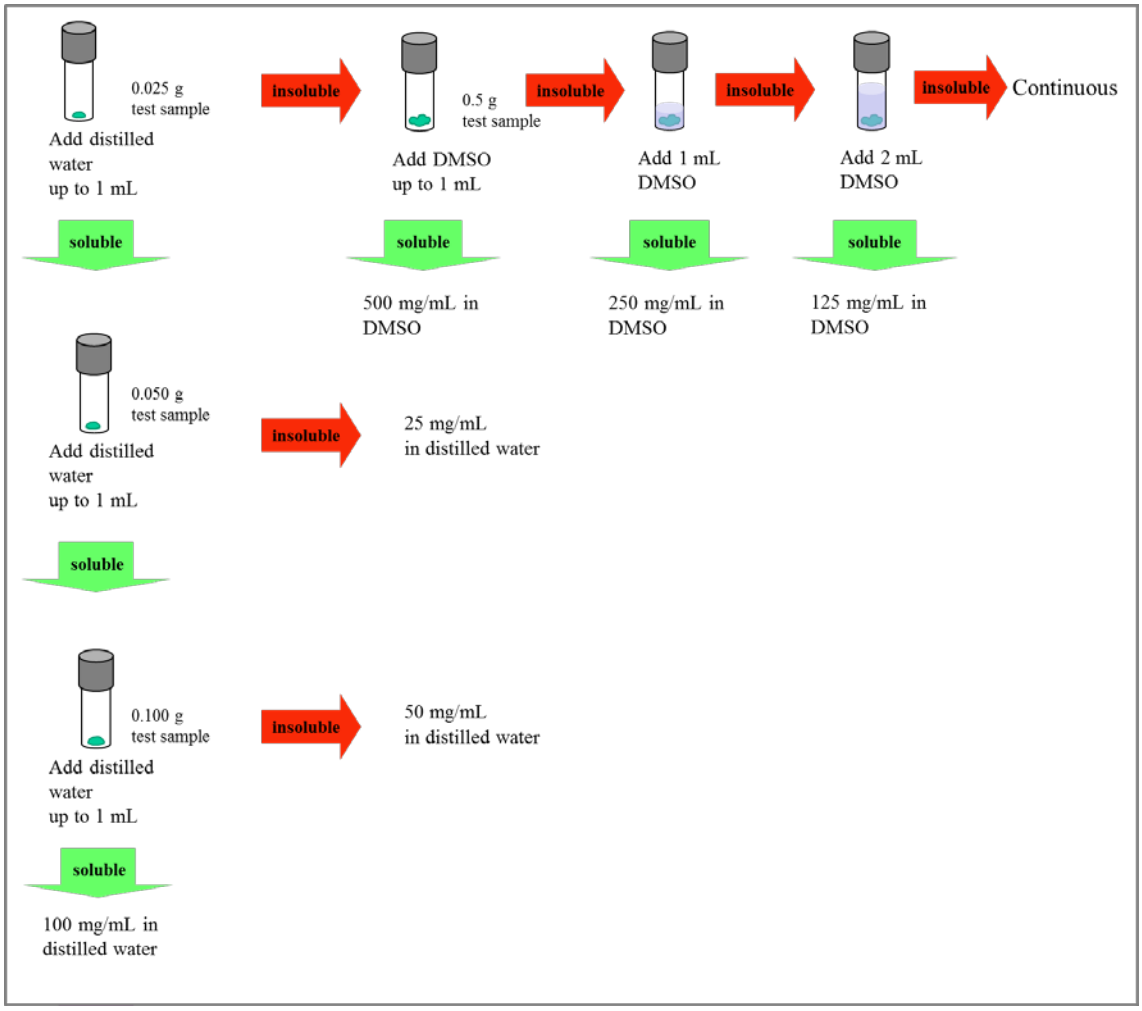


Fig. 12. Dissolution by vehicle

8-2-8. Dilution of chemicals

For water soluble chemicals, 11 serial dilutions were conducted using B medium, diluting by a factor of 2, in the 1st experiment. In the 2nd, 3rd, or 4th experiment, 11 serial dilutions were conducted, diluting by a factor of 1.5. For water insoluble chemicals, 11 serial dilutions were conducted using DMSO as the solvent, diluting by a factor of 2 in

the 1st experiment and by a factor of 1.5 in the 2nd, 3rd, and 4th experiments. The diluted chemicals are added to 2H4 cells in a 96 well plate. After one-hour incubation at 37°C in a 5% CO_2 incubator, 2H4 cells are added 10 µL of PMA/Io solution and incubated again at 37°C in a 5% CO_2 incubator for 6 hours.

8-2-9. Measurements

After incubation with the chemical and PMA/Io for 6 h at 37°C in a 5% CO_2 incubator, 100 µL of pre-warmed Tripluc is added to each well in the plate containing reference samples using a pipetman and the plate is shaken for 10 min at room temperature (about 25°C) using a plate shaker. Surface bubbles are removed if present and bioluminescence in each well is measured using a microplate-type luminometer with a multi-color detection system (Phelios; Atto Co., Tokyo, Japan) for 3 sec each in the absence (F0) and presence (F1, F2) of the optical filter. The F0, F1 and F2 data (values are expressed as counts) are processed using an Excel-based data sheet (Appendix 10). SLG-LA, SLO-LA and SLR-LA are calculated for each well based on the algorithm to calculate SLG-LA, SLO-LA and SLR-LA from the raw luminescence data reported previously (Nakajima et al., 2005; Noguchi et al., 2008). In addition to being used to calculate SLG-LA, SLO-LA and SLR-LA, this data sheet can automatically generate final graphs showing the correlation between %suppression and the concentration of chemicals, and between II-SLR-LA and the concentration of the chemical.

8-2-10. Luminometer apparatus

Multi-color detection systems such as microplate-type luminometers are available and include Phelios (ATTO, Tokyo, Japan), Tristan 941 (Berthold, Bad Wildbad, Germany), and the ARVO series (PerkinElmer, Waltham, MA). The luminometer detectors must have high sensitivity and low background noise and are usually equipped with optical filters, such as sharp-cut (long-pass) filters and band-pass filters. The transmission coefficients of these filters for each bio-luminescence signal color must be calibrated prior to all experiments following the manufacturer's

recommended protocol because the transmittance of the optical filter or the sensitivity of the detector are dependent on the measurement conditions.

8-2-11. Positive control

In each experimental set, dexamethasone and cyclosporine A are used as positive controls.

8-2-12. Calculation and definition of parameters for the IL-2 Luc assay

In the IL-2 Luc assay, the lead laboratory defined nIL2LA to represent IL-2 promoter activity by the SLG luciferase activity (IL2LA) normalized by SLR luciferase activity (GAPLA). The suppression index of GAPLA (Inh-GAPLA) was obtained by dividing GAPLA of 2H4 treated with chemicals with GAPLA of non-treated 2H4. % suppression reflects the effect of chemicals on IL-2 promoter. (Table 10).

Table 10. Abbreviations used in the 2H4 luciferase assay protocol

Parameter	Definition
IL2LA	Luciferase activity of stable luciferase green (Under the control of IL-2 promoter)
IFNLA	Luciferase activity of stable luciferase orange (Under the control of IFN-γ promoter)
GAPLA	Luciferase activity of stable luciferase red (Under the control of GAPDH promoter)
Normalized IL2LA (nIL2LA)	= (IL2LA) / (GAPLA)
Normalized IFNLA (nIFNLA)	= (IFNLA) / (GAPLA)
Inhibition index of GAPLA (Inh-GAPLA)	= (GAPLA of 2H4 treated with chemicals) / (GAPLA of untreated 2H4) (The cytotoxic effect of chemicals)
% suppression	= (1-(nIL2LA of 2H4 treated with chemicals) / (nIL2LA of non-treated 2H4)) x 100 (The effect of chemicals on IL-2 promoter)

8-2-13. Acceptance criteria

The following acceptance criteria should be satisfied when using the IL-2 Luc Assay method.

If Fold induction of nIFNLA of PMA/Io wells without chemicals (=(nIFNLA of 2H4 cells treated with PMA/Ionomycin) / (nIFNLA of non-treated 2H4 cells)) demonstrate less than 3.0, the results obtained from the plate containing the control wells should be rejected.

8-2-14. Prediction model

The experiments are repeated until 2 consistent suppressive (or stimulatory) results or 2 consistent “no effect results” are obtained. When 2 consistent results are obtained, the chemicals are judged as indicated by the obtained consistent results.

An immunotoxicant is identified by the mean of %suppression and its 95% simultaneous confidence interval.

In each experiment, when the chemicals clearly meet the following 3 criteria, they are judged as suppressive or stimulatory. Otherwise, they are judged as ‘no effect’ chemicals.

1. The mean of %suppression is ≥35 (suppressive) or ≤-35 (stimulatory) with statistical significance. The statistical significance is judged by its 95% confidence interval.
2. The result shows 2 or more consecutive statistically significant positive (negative) data points or 1 statistically significant positive (negative) data point with a trend in which at least 3 consecutive data points increase (decrease) in a dose-dependent manner. In the latter case, the trend can cross 0, as long as only 1 data point shows the opposite effect without statistical significance.
3. The results are judged using only data obtained at the concentration at which I.I.-SLR-LA is ≥0.05.

8-3. Data collection

8-3-1. Operating procedure

The detail of operating procedure in this assay is describe to the protocol version 011E. The version of protocols was updated during the validation studies, but for the operating procedure, the descriptions of operating procedure described in these protocols are same through the 2 validation studies.

8-3-2. Chemicals

For phase I study, in which the main aim was to evaluate intra- and inter-laboratory reliability, a total of 15 coded chemicals, for 3 rounds of 5 chemicals, were distributed to all the 3 laboratories. Because the different code between rounds was used, the technician

in each laboratory did not identify the same chemicals. For the phase II study, in which the main aim of phase I was to evaluate inter-laboratory reliability, 20 coded chemicals were distributed.

In this document the codes for the chemicals were re-coded. To indicate the round, the suffix is used such like P101_R1 for the first chemical of the first round in Phase I study: P1 means Phase I; 01 means the first chemical; _R1 means first round.

The Table 11 shows the chemical coded through this document.

Table 11. The chemical coded through this document

Phase	Chemical code	Lab A	Lab B	Lab C
I	P101_R1, P101_R2, P101_R3, P102_R1, P102_R2, P102_R3, P103_R1, P103_R2, P103_R3, P104_R1, P104_R2, P104_R3, P105_R1, P105_R2, P105_R3	3 rounds	3 rounds	3 rounds
II	P201, P202, P203, P204, P205, P206, P207, P208, P209, P210, P211, P212, P213, P214, P215, P216, P217, P218, P219, P220	1 round	1 round	1 round

8-3-3. Data handling

The developed Excel data sheet for this study was distributed to the laboratories. We had received data files from the 3 laboratories.

From JaCVAM we received files listed the chemical codes for the distributed 5 chemicals for the phase I study, and 20 chemicals for the phase II study.

For the data analysis, these files were combined and some datasets were constructed for the analysis. The SAS ver. 9.4 and Microsoft Excel was used for the data analysis described in this report.

Since the Excel data sheet is able to display a concentration-response plot for %suppression with its 95% confidence interval, we were able to judge “Suppressive”,

“Stimulatory” or “Negative” for each experiment by seeing the plot.

8-3-4. Index from each experiment and decision criteria for judgment

The j-th repetition (j = 1 to 4) of the i-th concentration (j = 0 to 11) is measured for IL2LA and GAPLA respectively. The normalized IL2LA is referred as nIL2LA, and is defined as

$nIL2LA_{ij} = IL2LA_{ij} / GAPLA_{ij.}$

This is the basic unit of measurement in this assay.

8-3-4-1. %suppression

The %suppression is an index for the averaged nIL2LA for the repetition on the i-th concentration compared with it on the 0 concentration, it is the primary measure of this assay. The %suppression is able to write by the following formula,

$$\%\ \text{suppression}_{\text{i}} = \left\{1 - \frac{\left(\frac{1}{4}\right)\sum_{\text{i}} \text{nIL2LA}_{\text{ij}}}{\left(\frac{1}{4}\right)\sum_{\text{i}} \text{nIL2LA}_{\text{0j}}}\right\} \times 100 \qquad (1)$$

The lead laboratory has proposed that ±35 of the value suggests suppressive and stimlatory for a tested chemical. This value is based on the investigation of the historical data of the lead laboratory. Data management team followed to use the value through all the phase of present validation study.

The primary outcome measure, % suppression, is basically the ratio of 2 arithmetic means of nIL2LA as shown in equation (1). The 95% confidence interval (95% CI) of the % suppression for the i-th concentration can be estimated.

The lower limit of the 95% CI above 0 is interpreted as that the nIL2LA with the i-th concentration is statistical-significantly greater than it with the 0-concentration, whereas the upper limit of the 95% CI blow 0 is interpreted as that the nIL2LA with the i-th concentration is statistical-significantly lesser than it with the 0-concentration.

There are several ways to construct the 95% CI. We used the method kwon as the

Delta method in this study. This 95% confidence interval theorem is obtained from the following formula.

$$\%\text{suppression} \pm 100 \times \left\{ z_{0.975} \times \sqrt{\frac{\text{sd}_\text{i}^2}{\text{mean}_0^2} + \frac{\text{mean}_\text{i}^2 \times \text{sd}_0^2}{\text{mean}_0^4}} \right\},$$

where mean_i is the mean of nIL2LA at the i-th concentration, mean_0 is the mean of nIL2LA at 0 concentration, sd_i is the standard deviation of nIL2LA at the i-th concentration and sd_0 is the standard deviation of nIL2LA at 0 concentration. $z_{0.975}$ is 97.5 percentile of the standard normal distribution.

8-3-4-2. Inh-GAPLA

The Inh-GAPLA is a ratio of the averaged GAPLA for the repetition of the i-th concentration compared with it of the 0 concentration, and this is written by

$$\text{Inh} - \text{GAPLA}_\text{i} = \{(1/4)\text{x} \sum_\text{j} \text{GAP LA}_\text{ij}\} / \{(1/4)\text{x} \sum_\text{j} \text{GAP LA}_\text{0j}\}$$

Since the GAPLA is the denominator of the nIL2LA, the extremely smaller value of this is considered to cause the large variation of the nIL2LA. Therefore, the i-th %suppression value with extremely smaller value of the Inh-GAPLA might be poor precision.

8-3-4-3. Judgment for “Suppressive”, “Stimulatory” or “No effect” in each experiment

In each experiment, when the following 3 criteria are satisfied, they are judged as “suppresive” or “stimulatory”. Otherwise, they are judged as no effect chemicals.

1. % suppression is ≥ 35 (suppressive) or ≤ -35 (stimulatory) at any dose and statistically significant.
2. The result shows two or more consecutive statistically significant positive (negative) data or one statistically significant positive (negative) data with a trend in which at least 3 consecutive data increase (decrease) in a dose dependent manner. In the latter case, the trend can cross 0, as long as only one data point shows the opposite effect without statistical significance.

3. The results are judged using only data obtained in the concentration at which Inh-GAPLA is $\geq$ 0.05

For 1, 2, the statistically significant is judged by the lower limit of 95% confidence interval of %suppression is over 0 or the upper limit of it is under 0.

8-3-4-4. Final judgment for "Suppressive" "Stimulatory" or "No effect" using this assay

In this assay, "Suppressive", "Stimulatory" or "No effect" is defined as in case that the 2 same judgments were found in a set of experiments.

8-3-5. Reliability

8-3-5-1. Within-laboratory reproducibility for 5 common chemicals

Within-laboratory reproducibility was determined by whether or not tables of 3 sets for the final judgment for each chemical by each laboratory were concordant. The concordance rate was then calculated as a proportion of the concordance of each laboratory.

The concordance rate for within-laboratory reproducibility was based on the results of 3 sets.

To summarize, the concordance rate for within-laboratory reproducibility from the 3 laboratories were used to calculate the averaged concordance rate.

8-3-5-2 Between-laboratory reproducibility

Between-laboratory reproducibility was determined using the results from the final judgment from the 3 laboratories for 25 chemicals, this is, 5 chemicals in Phase I study and 20 chemicals in Phase II study. These judgements were tabulated, then the concordance rate was calculated as a proportion of the concordance in each laboratory.

To summarize, the concordance rate for between-laboratory reproducibility from the 3 laboratories were used to calculate the averaged concordance rate.

.

8-3-6. Predictivity

In the evaluation of predictivity, we did not distinguish suppression and stimulation, because both of these indicate modulation of immune function. Then, we dealt as “Positive (P)” in case of “suppression” or “stimulation”, and “No effect (N)” in case of no significant effects for each chemical judgement.

The concordance, sensitivity and specificity were estimated as the indexes of predictivity. These indexes were estimated using the frequency results obtained from the 2 by 2 contingency table for T cell targeting. The definitions of these indexes are summarized in Table 12 below. This calculation was based on the results decided by a majority for the between-laboratory results for each chemical.

Table 12. Definition of the concordance, sensitivity and specificity

Judgment from the IL-2 Luc assay	Chemical category		Total
	Positive	Negative	
Positive	a	b	a+b
Negative	c	d	c+d
Total	a+c	b+d	N

Sensitivity = 100 × a / (a+c)

Specificity = 100 × d / (b+d)

Accuracy = 100 × (a+d) / N

8-4. Quality assurance

Assays and quality assurance were carried out in the spirit of GLP, although not all the participating laboratories routinely worked under GLP certification. The participating laboratories conducted the experiments in accordance with the protocol provided by the VMT. All raw data and data analysis sheets were pre-checked for quality by each laboratory and then were reviewed by the VMT quality assurance team.

The results accurately reflect the raw data.

9. Results

We conducted Phase I and II studies in this validation. The assay procedure and criteria used to judge immunotoxicants in the validation studies are summarized in Fig. 13.

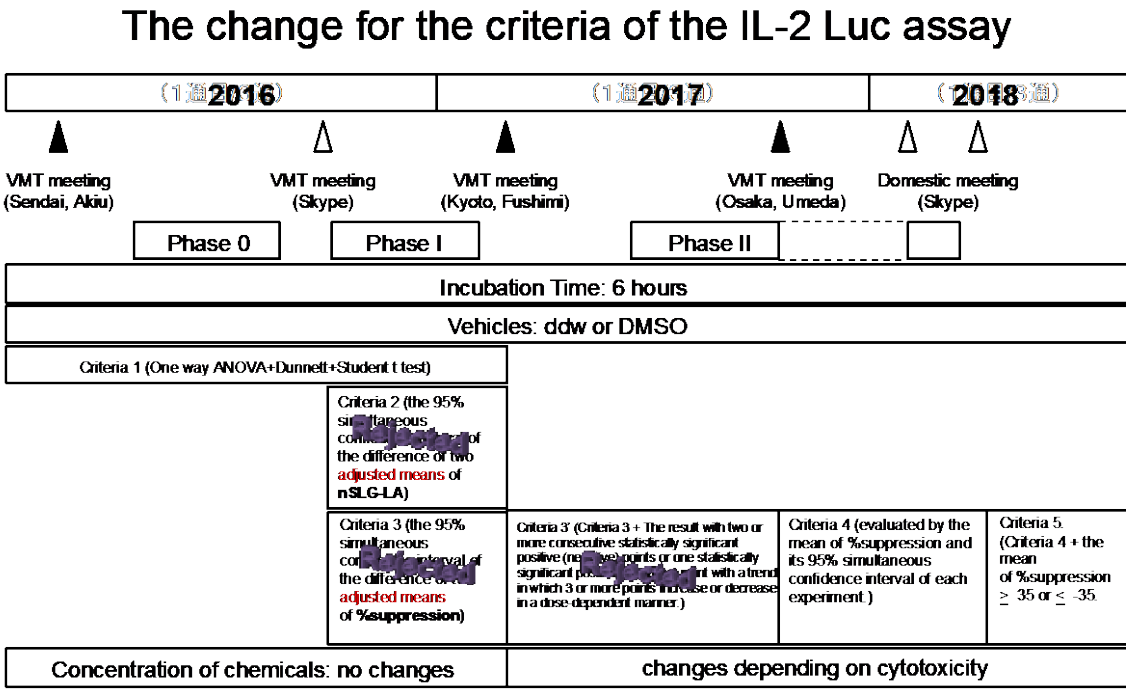


Fig. 13. The modification of the protocols of the IL-2 Luc assay.

9-1. The final criteria

9-1-1. Acceptance criteria

The following acceptance criteria should be satisfied when using the MITA method. In each time of the experiments, a control experiment examining nIL2LA of 2H4 cells treated with PMA/Io and nIL2LA of non-treated 2H4 cells must be conducted. Then, the fold induction of nIL2LA of PMA/Ionomycin wells without chemicals (= (nIL2LA of 2H4 cells treated with PMA/Ionomycin)/(nIL2LA of non-treated 2H4 cells)) is calculated. If the fold induction is less than 3.0, the results obtained from these experiments should be rejected.

9-1-2. Prediction model

The experiments are repeated until 2 consistent suppressive (or stimulatory) results or 2 consistent “no effect results” are obtained. When 2 consistent results are obtained, the chemicals are judged as indicated by the obtained consistent results.

An immunotoxicant is identified by the %suppression and its 95% simultaneous confidence interval.

In each experiment, when the chemicals clearly meet the following 3 criteria, they are judged as suppressive or stimulatory. Otherwise, they are judged as ‘no effect’ chemicals.

1. The mean of % suppression is ≥35 (suppressive) or ≤-35 (stimulatory) with statistical significance. The statistical significance is judged by its 95% confidence interval.
2. The result shows 2 or more consecutive statistically significant positive (negative) data points or 1 statistically significant positive (negative) data point with a trend in which at least 3 consecutive data points increase (decrease) in a dose-dependent manner. In the latter case, the trend can cross 0, as long as only 1 data point shows the opposite effect without statistical significance.
3. The results are judged using only data obtained at the concentration at which Inh-GAPLA is ≥0.05.

9-1-3. Predictivity

To determine the performance of the IL-2 Luc assay, it is crucial to understand the immunotoxicological characteristics of the chemicals used in the study. Since the IL-2 Luc assay focuses on the effects of chemicals on IL-2 transcription by T cells, we attempted to classify the chemicals into two categories: (i) immunotoxic chemicals which target T cells (TTCs), which include chemicals that directly affect T cell viability, T cell proliferation or T cell function and (ii) others (NTTCs), which include chemicals that do not directly affect T cell viability, T cell proliferation or T cell function. In this assay, to define TTCs, we first surveyed the literature and collected the following six findings regarding each of the chemicals proposed for use in the study in Table 13.

Table 13. The immunotoxicological data obtained from the literature.

Endopoints	Information
Endpoint 1	Decreased thymus weight
Endpoint 2	Increased or decreased IL-2, IFN-γ, IL-4 or other T cell-specific cytokine mRNA expression or protein production by T cells in *ex vivo.*
Endpoint 3	Increased or decreased IL-2, IFN-γ, IL-4 or other T cell-specific cytokine mRNA expression or protein production by T cells *in vitro.*
Endpoint 4	Suppressed T cell proliferation
Endpoint 5	Suppressed cytotoxic T cell response
Endpoint 6	The NTP data clearly indicate that one of the immunotoxic mechanism of chemicals are attributed to its effect on T cells.

Then, according to the rationale for classifying immunotoxic chemicals reported by Luster et al (Luster et al., 1992b), we defined TTCs as chemicals that satisfy one of the following criteria and then, made the reference data on immunotoxicity of chemicals in Table 14.

Table 14. The criteria to classify immunotoxic chemicals by affecting T cells.

Criteria	Definition
Criterion 1	Decreased thymus weight with additional one or more findings among endpoints 2 to 5
Criterion 2	Increased or decreased mRNA expression or protein production in one or more cytokines in Endpoints 2 or 3 in multiple reports
Criterion 3	Increased or decreased mRNA expression or protein production in two or more cytokines in Endpoints 2 or 3
Criterion 4	The presence of data suggesting that one of the immunotoxic mechanisms of the chemical was attributed to an effect on T cells in Endpoint 6

Then, by comparing the results of the IL-2 Luc assay (positive or no effect) with the classification of the chemicals (TTC or NTTC), we calculated the accuracy, sensitivity and specificity of the IL-2 Luc assay in the validation study.

To classify 25 chemicals used in the Phase I and II studies, we used the chemical information kindly provided by the National Toxicology Program (NTP). The immunotoxic characteristics of each chemical are shown in Appendix 7. The summarized data of the NTP data and the data collected by the VMT member are shown in Appendix 19. The list of references is in Appendix 8. As already described, IL-2 exerts pleiotropic actions on CD4+ T cell differentiation via its modulation of cytokine receptor expression. Indeed, IL-2 promotes Th1 and Th2 differentiation, while it also drives Treg differentiation. Therefore, it suggests that the augmentation of IL-2

transcription can lead to either immunostimulation or immunosuppression depending on surrounding tissue environment *in vivo.* Therefore, in this assay, if chemicals were judged as either stimulation or suppression, they were both considered as positive (P) and if not, they were judged as negative (N).

9-2. Phase 0 study (for technical transfer)

The preliminary test trial (Phase 0) was performed by the participating laboratories following explicit explanations of the Multi-ImmunoTox Assay protocol Ver. 008.1E by the lead laboratory, Tohoku University. Three laboratories participated in the Phase 0 study of the IL-2 Luc assay using 5 open labeled chemicals, 2-aminoantracene, citral, chloroquine, dexamethasone, methyl mercuric chloride and conducted 1 set (3 experiments) for each chemical. Most response patterns for the 5 chemicals were similar among the 3 laboratories except for 2 early experiments conducted by the naïve laboratory. Based on these results, VMT judged that technical and protocol transfer of the IL-2 Luc assay is acceptable.

After the Phase 0 study, we amended the protocol as follows:

- We changed the speed of centrifugation of the cells, and the preparation method for the selection antibiotics and PMA/Io.
- We set nIFNLA >3 as an acceptance criterion.
- Because nIL2LA is dependent on the properties of the specific luminometer used, we expressed the results of the data by %suppression, which is determined by dividing nIL2LA of the chemically treated cells by nIL2LA of the vehicle-treated cells.
- Volatile chemicals were to be sealed.
- We determined the criteria to judge chemicals from a statistical standpoint (Criteria 2).

9-3. Phase I study (for within and between-laboratory reproducibility)

9-3-1. Test conditions

A total of 5 coded chemicals (4 T cell targeting and 1 non-T cell targeting) were evaluated by 3 experimental sets in the Phase I study based on the Multi-ImmunoTox Assay protocol Ver. 011E.

In each experimental set, 3 or more experiments were conducted for each chemical.

Chemicals that satisfied criteria 5 were judged as positive. Chemicals that provided 2 positive results were judged as immunotoxicants in Tables 15 and 16.

9-3-2. Within-laboratory variation assessments in the Phase I study

Lab A	80.0% (4/5)
Lab B	100% (5/5)
Lab C	80.0% (4/5)
Average	86.7% (13/15)

9-3-3. Between-laboratory variation assessments in the Phase I study

Between-Lab reproducibility (Based on Majority)

80.0% (4/5)

9-3-4. Predictivity in the Phase I study (Based on Majority)

Accuracy of Lab A	80.0% (4/5)
Accuracy of Lab B	100% (5/5)
Accuracy of Lab C	100% (5/5)
Average	93.3% (14/15)

Table 15. Results of the Phase I study

Chemical	CAS	Set	Lab. A	Lab. B	Lab. C	Concordance	T cell targeting	Rationale
Dibutyl phthalate	84-74-2	1st	P	P	P	1	Yes	2, 3
		2nd	P	P	P			
		3rd	P	P	P			
Hydrocortisone	50-23-7	1st	P	P	P	0	Yes	1, 2
		2nd	N	P	P			
		3rd	N	P	N			
Lead(II) acetate	6080-56-4	1st	P	P	P	1	Yes	1, 2, 3
		2nd	P	P	P			
		3rd	P	P	P			
Nickel(II) sulfate	10101-97-0	1st	P	P	P	1	Yes	1, 2, 3
		2nd	P	P	P			
		3rd	P	P	P			
Zinc dimethyldithiocarbamate (DMDTC)	137-30-4	1st	N	N	N	1	No	
		2nd	N	N	N			
		3rd	N	N	N			
Within-laboratory reproducibility (%)			80.0 (4/5)	100 (5/5)	80.0 (4/5)			
			Average 86.7 (13/15)					
Between-laboratory reproducibility (%) (Based on Majority)						80 (4/5)		
Sensitivity (%) (Based on			75.0	100	100			

Majority)	(3/4)	(4/4)	(4/4)
	Average		
	91.7 (11/12)		
Specificity (%) (Based on Majority)	100 (1/1)	100 (1/1)	100 (1/1)
	100 (3/3)		
Accuracy (%) (Based on Majority)	80.0 (4/5)	100 (5/5)	100 (5/5)
	Average		
	93.3 (14/15)		

P: Positive, N : No effect

9-3-5. Contingency tables for the Phase I study

Table 16. Contingency tables for the Phase I study

Lab A		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	10	2	12
	-	0	3	3
Total		10	5	15

Sensitivity : 83.3% (10/12)

Specificity : 100% (3/3)

Accuracy : 86.7% (13/15)

Lab B		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	12	0	12
	-	0	3	3
Total		12	3	15

Sensitivity : 100% (12/12)

Specificity : 100% (3/3)

Accuracy : 100% (15/15)

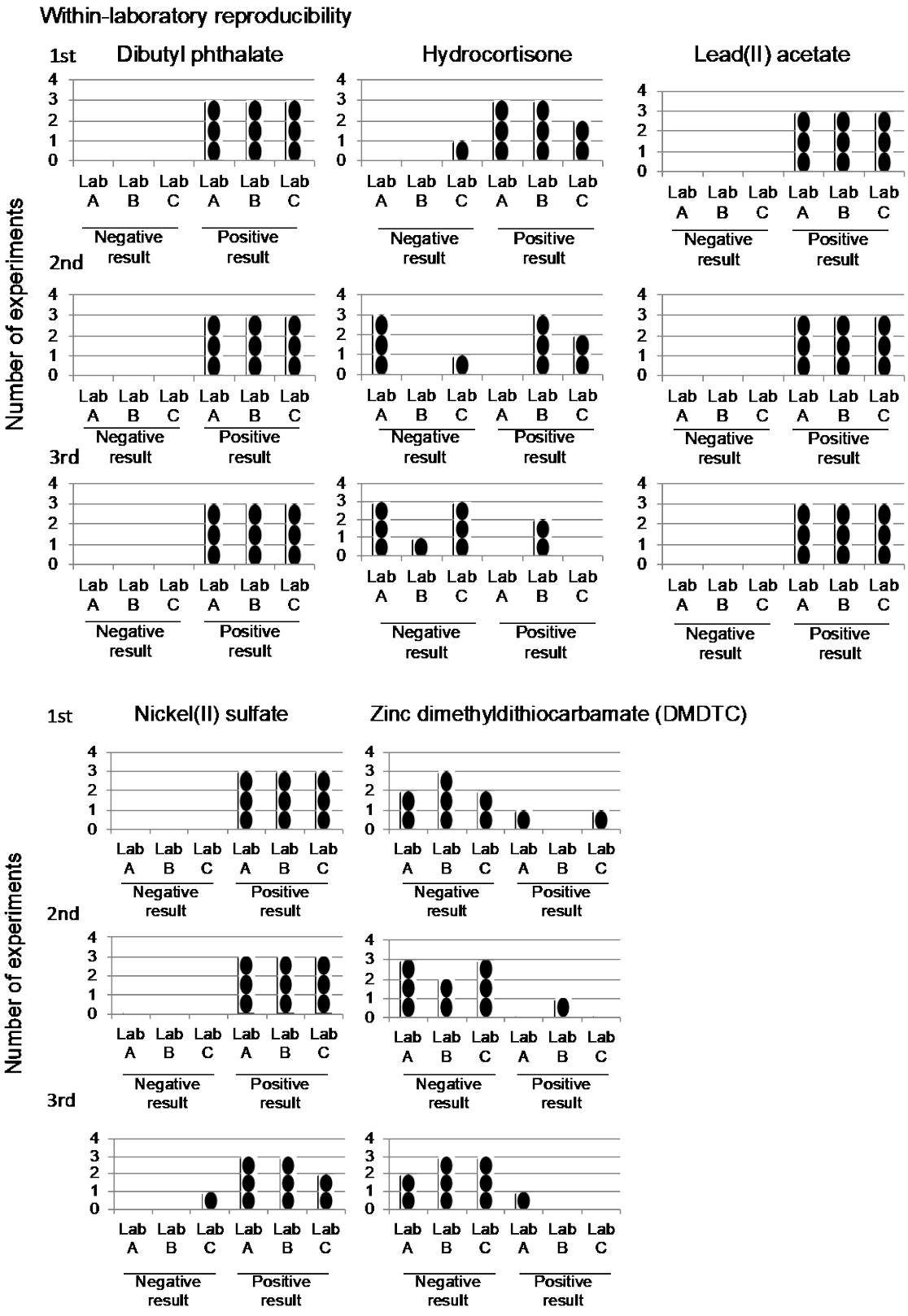
Lab C		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	11	1	12
	-	0	3	3
Total		11	4	15

Sensitivity : 91.7% (11/12)

Specificity : 100% (3/3)

Accuracy : 93.3% (14/15)

A graphical presentation of between- and within-laboratory variation in Phase I study is shown in Fig. 14.


Within-laboratory reproducibility
1st
2nd
3rd
Dibutyl phthalate
Hydrocortisone
Lead(II) acetate
Nickel(II) sulfate
Zinc dimethyldithiocarbamate (DMDTC)
Number of experiments
0
1
2
3
4
Lab A
Lab B
Lab C
Negative result
Positive result

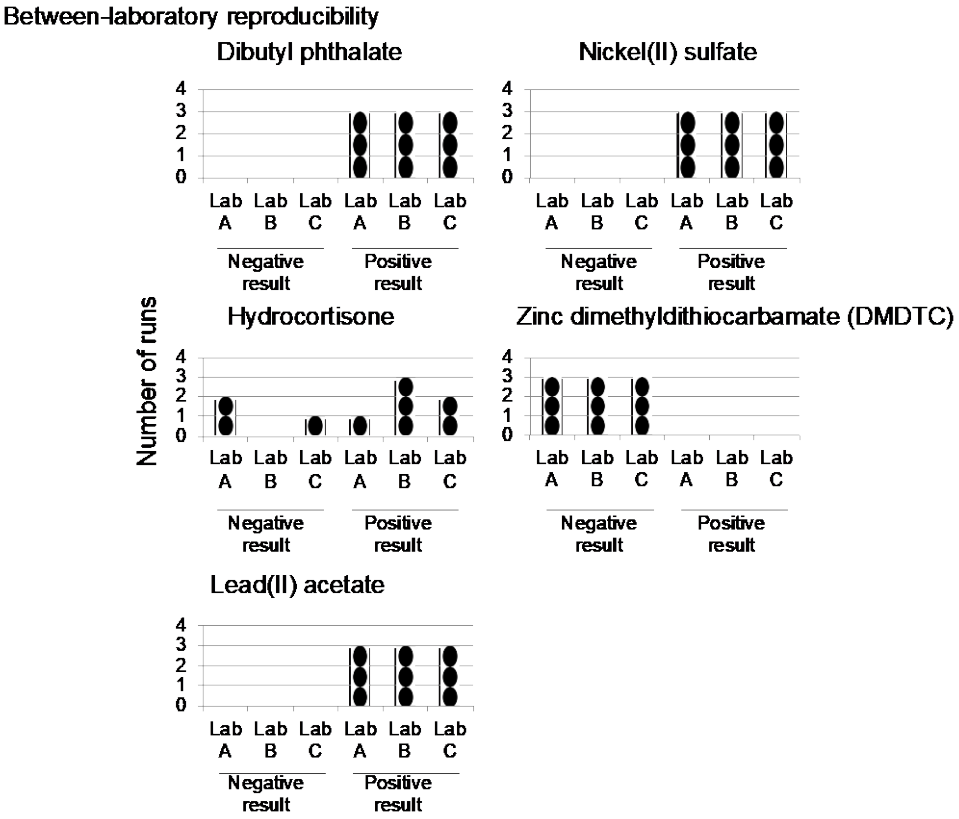


Fig. 14. Between- and within- laboratory variation assessments in the Phase I study

The Phase I study examined within-and between-laboratory reproducibilities using a total of 5 coded chemicals (4 T cell targeting and 1 non-T cell targeting) evaluated by 3 experimental sets based on the MITA protocol Ver. 008.5E. Closed circles represent the judgments for individual experiments for within-laboratory reproducibility or the judgments in individual experimental sets for between-laboratory reproducibility.

9-4. Phase II study (for between-laboratory reproducibility and predictivity)

9-4-1. Test conditions

The Phase II study for between-laboratory reproducibility and predictivity was conducted using a total of 20 coded chemicals (12 T cell targeting, 7 non-T cell targeting and 1 undetermined) and evaluated by 1 experiment set based on the Multi-ImmunoTox Assay protocol Ver. 011E in Tables 17 to 19.

9-4-2. Between-laboratory variation assessments in the Phase II study

Between-Lab reproducibility 80% (16/20)

9-4-3. Predictivity in the Phase II study

Accuracy of Lab A	73.7 (14/19)
Accuracy of Lab B	68.4% (13/19)
Accuracy of Lab C	68.4% (13/19)
Average	70.2% (40/57)

Table 17. Results of the Phase II study

Chemical	CAS	Lab.A	Lab.B	Lab.C	T cell targeting	Rationale
2,4-Diaminotoluene	95-80-7	N	N	N	No	
Benzo(a)pyrene	50-32-8	P	P	P	Yes	2), 3)
Cadmium chloride	10108-64-2	N	N	N	Yes	2), 3)
Dibromoacetic acid	631-64-1	P	P	N	Yes	1), 4)
Diethylstilbestol	56-53-1	P	P	P	Yes	1), 2), 4)
Diphenylhydantoin	630-93-3	N	N	N	Yes	2), 3), 4)
Ethylene dibromide	106-93-4	N	N	N	Yes	1)
Glycidol	556-52-5	P	P	P	No	
Indomethacin	53-86-1	P	P	P	Yes	3), 4)
Isonicotinic Acid Hydrazide	54-85-3	P	N	P	Yes	2)
Nitrobenzene	98-95-3	N	P	N	Undetermined	
Urethane, Ethyl carbamate	51-79-6	P	P	P	Yes	1)
Tributyltin chloride	1461-22-9	P	P	P	Yes	1)
Perfluorooctanoic acid	335-67-1	P	P	P	Yes	1)
Dichloracetic acid	79-43-6	P	P	P	Yes	2), 3)
Toluene	108-88-3	N	N	N	No	
Acetonitril	75-05-8	N	N	N	No	
Mannitol	69-65-8	N	N	N	No	
Vanadium pentoxide	1314-62-1	N	N	N	No	
o-Benzyl-p-chorolophenol	120-32-1	P	P	P	No	

Table 18. Reproducibility of the Phase II study

Between-laboratory reproducibility(%) 80 (16/20)			
Sensitivity (%)	75.0 (9/12)	66.7 (8/12)	66.7 (8/12)
Specificity (%)	71.4 (5/7)	71.4 (5/7)	71.4 (5/7)
Accuracy (%)	73.7 (14/19)	68.4 (13/19)	68.4 (13/19)

P: Positive, N : No effect

9-4-4. Contingency tables for the Phase II study

Table 19. Contingency tables for the Phase II study

Lab A		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	9	3	12
	-	2	5	7
Total		11	8	19

Sensitivity	75.0 (9/12)
Specificity	71.4 (5/7)
Accuracy	73.7 (14/19)

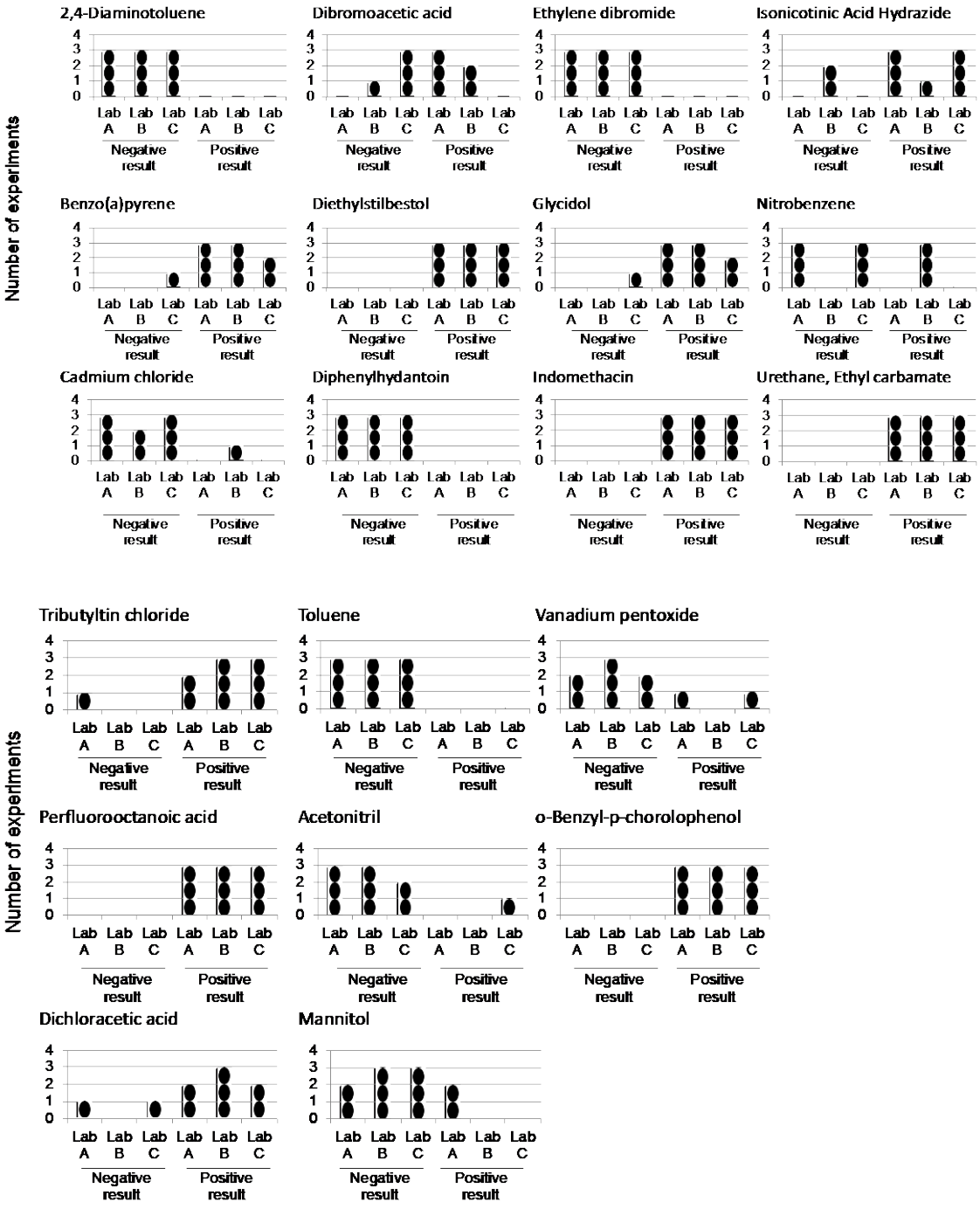
Lab B		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	8	4	12
	-	2	5	7
Total		10	9	19

Sensitivity	66.7 (8/12)
Specificity	71.4 (5/7)
Accuracy	68.4 (13/19)

Lab C		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	8	4	12
	-	2	5	7
Total		10	9	19

Sensitivity	66.7 (8/12)
Specificity	71.4 (5/7)
Accuracy	68.4 (13/19)

The graphical presentation of between- and within-laboratory variation in Phase II study is Fig 15.


Number of experiments
2,4-Diaminotoluene
Dibromoacetic acid
Ethylene dibromide
Isonicotinic Acid Hydrazide
Benzo(a)pyrene
Diethylstilbestol
Glycidol
Nitrobenzene
Cadmium chloride
Diphenylhydantoin
Indomethacin
Urethane, Ethyl carbamate
Tributyltin chloride
Toluene
Vanadium pentoxide
Perfluorooctanoic acid
Acetonitril
o-Benzyl-p-chorolophenol
Dichloracetic acid
Mannitol
0
1
2
3
4
Lab A
Lab B
Lab C
Negative result
Positive result

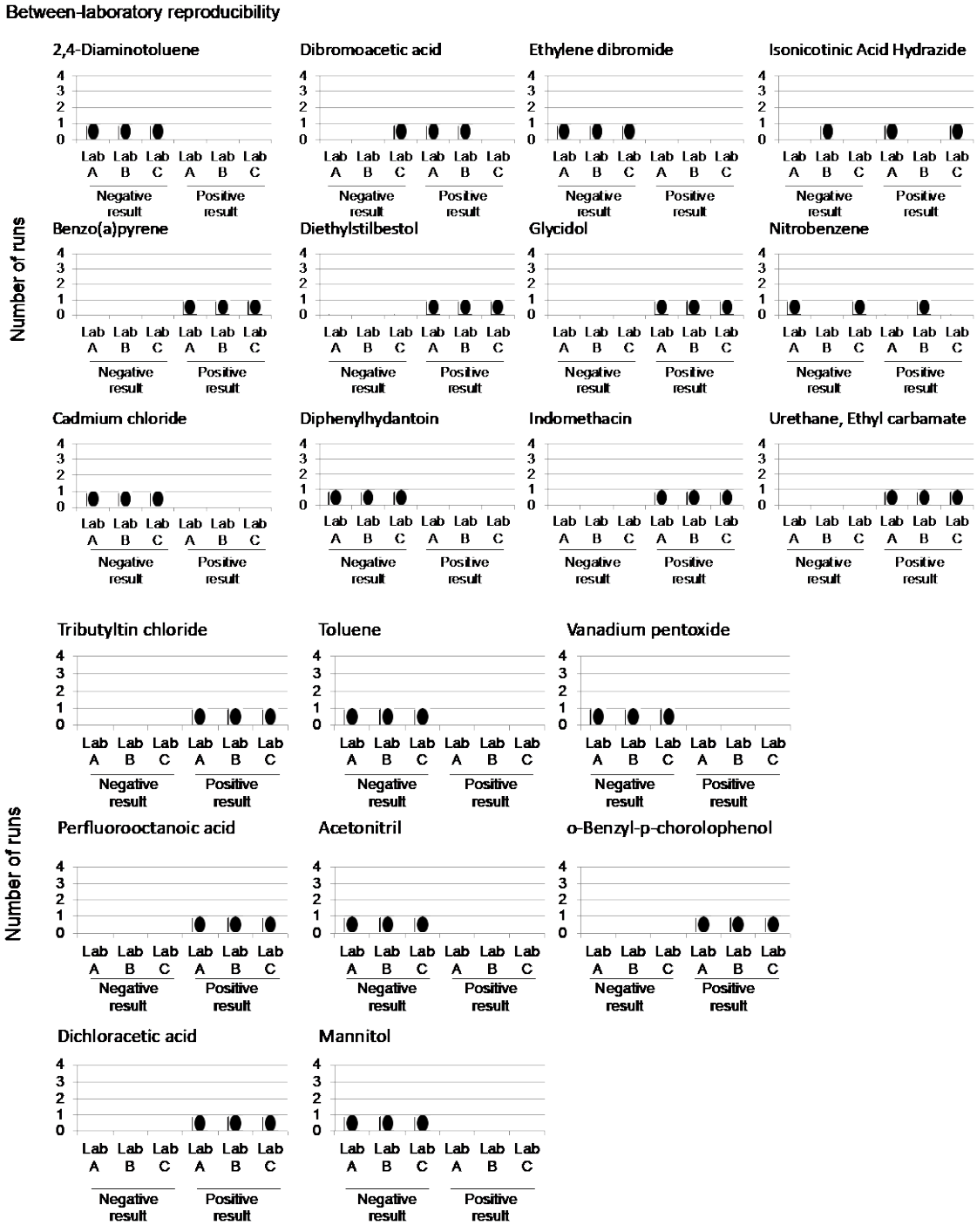


Fig. 15. Between variation assessments in the Phase II study

The Phase II study examined between-laboratory reproducibility using a total of 20 coded chemicals (13 T cell targeting, 6 non-T cell targeting and 1 undetermined) evaluated by 1 experiment sets based on Multi-ImmunoTox Assay protocol Ver. 011E. Closed circles represent the judgments for individual experiments in within-laboratory reproducibility or represent the judgments in individual experimental sets for between-laboratory reproducibility.

9-5. Quality assurance

9-5-1. 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 (Appendix 14). JaCVAM was responsible for coding and distributing the test chemicals for the validation study.

9-5-2. 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. 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.

9-5-3. Independent analysis by the biostatistician

All data sheets from the participating laboratories were collected and checked by Dr. Takashi Omori, Kobe univ., the independent biostatistician and JaCVAM. Dr. Omori and his colleagues summarized the data (Appendix 11) and the concentration-response plot for each experiment in phase I (Appendix 17) and phase II (Appendix 18).

9-5-4. Quality assurance by JaCVM

All the record sheets from the participating laboratories were also checked and JaCVAM (Appendix 13). 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:// http://jacvam-jp.check-xserver.jp/validation08-login.html). Testings performed as part of a validation study were carried out 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.

The culture of the cells, the preparation and application of test chemicals and data sheets were completed and the results accurately reflect the raw data. Unfortunately, the record sheets on the maintenance of measuring instruments had not collected before the validation study. JaCVAM considered these records had concerns on quality of data in the validation study. However, JaCVAM checked carefully all the results and judged all data within acceptable ranges.

At least, the reliability of measuring instruments would be checked by an independent organization before the validation study. JaCVAM recommend the validation management team the formal validation study participated with GLP laboratories will be done.

9-6. Combined results of the Phase I and II studies (for between- and within-laboratory reproducibility and predictive capacity)

9-6-1. Test conditions

The within- and between-laboratory reproducibilities, and the predictivity of the IL-2 Luc assay, were evaluated using all the results from Phases I and II in Tables 20 to 22.

9-6-2. Within- and between-laboratory variation assessments from the Phase I and II studies.

Between-Lab reproducibility 80% (20/25)

Within-Lab reproducibility Lab. A 80.0% (4/5)

Lab. B 100% (5/5)

Lab. C 80.0% (4/5)

Average 86.7% (13/15)

9-6-3. Predictivity in the Phases I and II studies

Accuracy of Lab. A 75.0% (18/24)

Accuracy of Lab. B 75.0% (18/24)

Accuracy of Lab. C 75.0% (18/24)

Average 75.0% (54/72)

Table 20. Combined results of the Phase I and II studies

Chemical	CAS	Lab.A	Lab.B	Lab.C	concordance	T cell targeting
Phase I						
Dibutyl phthalate	84-74-2	PPP	PPP	PPP	1	Yes
Hydrocortisone	50-23-7	PNN	PPP	PPN	0	Yes
Lead(II) acetate	6080-56-4	PPP	PPP	PPP	1	Yes
Nickel(II) sulfate	10101-97-0	PPP	PPP	PPP	1	Yes
Zinc dimethyldithiocarbamate (DMDTC)	137-30-4	NNN	NNN	NNN	1	No
Phase II						
2,4-Diaminotoluene	95-80-7	N	N	N	1	No
Benzo(a)pyrene	50-32-8	P	P	P	1	Yes
Cadmium chloride	10108-64-2	N	N	N	1	Yes
Dibromoacetic acid	631-64-1	P	P	N	0	Yes
Diethylstilbestol	56-53-1	P	P	P	1	Yes
Diphenylhydantoin	630-93-3	N	N	N	1	Yes
Ethylene dibromide	106-93-4	N	N	N	1	Yes
Glycidol	556-52-5	P	P	P	1	No
Indomethacin	53-86-1	P	P	P	1	Yes
Isonicotinic Acid Hydrazide	54-85-3	P	N	P	0	Yes
Nitrobenzene	98-95-3	N	S	N	0	Undetermined
Urethane, Ethyl	51-79-6	P	P	P	1	Yes

carbamate						
Tributyltin chloride	1461-22-9	P	P	P	1	Yes
Perfluorooctanoic acid	335-67-1	P	P	P	1	Yes
Dichloracetic acid	79-43-6	P	P	P	1	Yes
Toluene	108-88-3	N	N	N	1	No
Acetonitril	75-05-8	N	N	N	1	No
Mannitol	69-65-8	N	N	N	1	No
Vanadium pentoxide	1314-62-1	N	N	N	1	No
o-Benzyl-p-chorolophenol	120-32-1	P	P	P	1	No

Table 21 Reproducibility of the Phase I and II studies

Within-laboratory reproducibility (%)	80 (4/5)	100 (5/5)	80 (4/5)
	Average 86.7 (13/15)		
Between-laboratory reproducibility (%) (Based on majority for Phase I) 80 (20/25)			
Sensitivity (%)	75.0 (12/16)	75.0 (12/16)	75.0 (12/16)
	Average 75.0 (36/48)		
Specificity (%)	75.0 (6/8)	75.0 (6/8)	75.0 (6/8)
	Average 75.0 (18/24)		
Accuracy (%)	75.0 (18/24)	75.0 (18/24)	75.0 (18/24)
	Average 75.0 (54/72)		

P: Positive, N : No effect

9-6-4. Contingency tables for the Phase I and II studies

Table 22. Contingency tables for the Phase I and II studies

Lab A		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	12	4	16
	-	2	6	8
Total		14	10	24

Sensitivity : 75.0 % (12/16)

Specificity : 75.0 % (6/8)

Accuracy : 75.0 % (18/24)

Lab B		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	12	4	16
	-	2	6	8
Total		14	10	10

Sensitivity : 75.0 % (12/16)

Specificity : 75.0 % (6/8)

Accuracy : 75.0 % (18/24)

Lab C		IL-2 Luc assay		Total
		+	-	
T cell targeting	+	12	4	16
	-	2	6	8
Total		14	10	10

Sensitivity : 75.0 % (12/16)

Specificity : 75.0 % (6/8)

Accuracy : 75.0 % (18/24)

10. Discussion

10-1. Reliability

The IL-2 Luc assay is based on the modulation of PMA + ionomycin-induced luciferase activity in the IL-2 reporter cell line, 2H4. Therefore, it is crucial that 2H4 cells maintain their ability to induce luciferase activity following stimulation by PMA/Io. Before and during this validation study, the response of 2H4 cells to PMA/Io was carefully observed. We confirmed that a frozen stock of 2H4 cells can be cultured without losing luciferase activity for at least 16 weeks or 35 passages.

The culture of 2H4 cells is relatively simple and does not require the use of trypsin or EDTA because 2H4 cells do not adhere to the culture dishes. First, cells adjusted to the optimum concentration are seeded into each well of a 96-well culture plate. Then, chemicals at graded concentrations are added to the wells. After 6 h incubation, 100 μL of pre-warmed Tripluc is added to each of the 96 wells. The subsequent process is completely automated, except for calculating the results using the predesigned Excel spreadsheet. Therefore, the IL-2 Luc assay is a test method that can significantly reduce human error.

Moreover, the IL-2 Luc assay does not require the determination of cell viability after chemical treatment. 2H4 cells can present IL-2 promoter activity as well as promoter activity of GAPDH, a well-known housekeeping gene; therefore, information regarding the effects of the chemical on both IL-2 induction and cell viability is obtained in each experiment. Furthermore, a single experiment takes only 8 h, including the time required for chemical preparation and cell plating, making the IL-2 Luc assay a true high-throughput method.

10-2. Between- and within-laboratory reproducibility

We examined within-laboratory reproducibility in the Phase I study. Lab A, Lab B, and Lab C demonstrated 80%, 100%, and 80% reproducibility, respectively. On the other hand, the between-laboratory reproducibility of Lab A, Lab B, and Lab C demonstrated

80% in the combined data of the Phase I and Phase II studies. These results satisfied the acceptance criteria for the validation study with a within-laboratory reproducibility of at least 80% and a between-laboratory reproducibility of at least 80%.

10-3. Predictivity

10-3-1. Rationale to determine the predictivity of the IL-2 Luc assay by the concordance between positive effects and the immunotoxic effects targeting T cell response

Reference data showing which chemicals are immunotoxic are essential for determining the performance of the IL-2 Luc assay. However, such reference data are lacking for most chemicals and thus we attempted to create reference data for the chemicals used in this study. Although there is no gold standard to date for classifying immunotoxic chemicals, Luster et al. (Luster et al., 1992b) proposed a rationale for immunotoxic classification, when they presented a screening battery using a 'tier' approach for detecting potential immunotoxic compounds in mice (Luster, 1998). Their proposal was that a positive reference chemical would either produce a significant dose-response effect in the immune test or significantly alter two or more immune test results at the highest dose of the chemical tested. They classified chemicals based on the 1results obtained in 12 immune tests according to this rationale and found a significant correlation between the judgment of immunotoxic chemicals and host resistance (Luster et al., 1993). Therefore, we used this rationale and classified chemicals based on the published previously immunotoxicological information for each chemical.

When immunotoxic information of chemical is collected from the literature, however, most of the published data are not focusing on whether immunotoxicity of chemicals is caused either by their direct effects on T cell or not. To overcome this problem, in this study, the predictivity was evaluated by the criteria whether chemicals affect T cell functions, namely T cell targeting, or not. To determine T cell targeting chemicals (TTCs), we defined the criteria described in 9-1-3.

10-3-2. The predictivity of the Phase I and Phase II studies

To classify 25 chemicals used in the Phase I and II studies, we used the chemical information kindly provided by the National Toxicology Program (NTP) and those collected by the VMT members. The immunotoxic characteristics of each chemical are shown in Appendix 7 and their summarized data are shown in the Appendix 19. Based on the criteria, the 25 chemicals were classified into 16 TTCs, 8NTTCs, and 1 unclassified chemicals that could not be classified because of insufficient data. According to this classification, the sensitivities of the assays as conducted by Lab A, Lab B, Lab C, and their average in the combined data of the Phase I and II studies are 75.0%, 75.0%, 75.0% and 75.0%, respectively. The specificities of the assays as conducted by Lab A, Lab B, Lab C, and their average are 75.0%, 75.0%, 75.0%, and 75.0%, respectively. The accuracies of the assays conducted by Lab A, Lab B, Lab C, and their average are 75.0%, 75.0%, 75.0%, and 75.0%, respectively.

10-4. IL-2 Luc assay data set for 60 chemicals

Based on the Multi-ImmunoTox assay protocol Ver. 011E and the Criteria 5, the lead laboratory reevaluated the data of 60 chemicals reported previously (Kimura et al. 2018) (Table 23). These 60 chemicals were also classified by the criteria described in 9-1-3. The classification of chemicals and their immunotoxic information were summarized in the Appendix 20. The list of references is in the Appendix 9. There were 34 TTCs, 6 NTTCs, and 20 chemicals that were either those without any immunotoxic information or with insufficient information. Similar to the classification by the criteria used in our published paper (Kimura et al., 2018), TAC, CyA, and Dex significantly suppressed IL-2 luciferase activity (IL-2LA), although the average LOEL of TAC and CyA was significantly lower that of DEX. The off-label immunosuppressive drugs, chloroquine, minocycline, and dapsone significantly suppressed IL-2LA. Anti-cancer drugs, actinomycin D and cisplatin also significantly suppressed IL-2LA. In addition, azathioprine and colchicine were demonstrated to suppress IL-2LA by the Criteria 5.

Again, the suppressive effects on the IL-2LA was not demonstrated by some of immunosuppressants the mechanism of which is inhibition of DNA synthesis or anti-proliferative effects on T cells, such as mitomycin C, cyclophosphamide, methotrexate or mizoribine by the Criteria 5.

If we calculated the predictivity of 60 chemicals evaluated by the IL-2 Luc assay based on the classification of chemicals defined in 9-1-3, the sensitivity, specificity and accuracy (predictivity) are 82.4% (28/34), 83.3% (5/6), and 82.5% (33/40), respectively.

Table 23. Data set of the IL-2 Luc assay based on Criteria 5.

Chemical name	Immunotoxicity classification		IL-2 Luc assay	Ave.LOEL(35%)	Ave.LOEL(-35%)
	Classification	Rationale*			
FK506	TTC	1,3	P	0.0002	
Cyclosporine A	TTC	1,3	P	0.0041	
Actinomycin D	TTC	3	P	0.0156	
Digoxin	TTC	2, 3	P	0.0686	
Colchicine	TTC	2, 3	P	0.2743	
FR167653	Undetermined	2, 3	P	1.3021	
Benzethonium chloride	Undetermined	1	P	1.6276	
Mercuric chloride	TTC	1,3	P	1.9531	
Chlorpromazine	TTC	1,3	P	1.9531	
Amphotericin B	Undetermined	1	P	2.6042	
Dibutyl phthalate	TTC	3	P	2.6042	
2-Aminoanthracene	Undetermined		P	5.8594	
Formaldehyde	TTC	2,3	P	7.8125	
Pyrimethamine	Undetermined		P	7.8125	
Isophorone diisocyanate	Undetermined		P	15.6250	
Cisplatin	TTC	1,2,3	P	16.9271	
Cobalt chloride	TTC	1, 3	P	16.9271	
Chloroquine	TTC	1,3	P	17.8326	
Minocycline	TTC	3	P	18.5185	
Mitomycin C	Undetermined		P	20.0000	
Hydrogen peroxide	TTC	3	P	23.4375	
Citral	Undetermined	1	P	25.0000	
Dexamethasone	TTC	1,3	P	41.1692	
Pentamidine isethionate	TTC	3	P	52.0833	
Lead(II)acetate	TTC	1, 3	P	57.2917	
Azathioprine	TTC	1, 2, 3	P	58.4778	
Diesel exhaust particle	TTC	1, 3	P	62.5000	
Sodium dodecyl sulfate	TTC	3	P	62.5000	
Dapsone	TTC	3	P	72.9167	
Nitrofurazone	NTTC		P	83.3333	
p-Nitroaniline	TTC	1,3	P	83.3333	
Sulfasalazine	TTC	1,3	P	92.9444	
Aluminium chloride	TTC	1,3	P	104.1667	
Nickel sulfate	TTC	1, 3	P	104.1667	
Hydrocortisone	TTC	1,3	P	125	
Diethanolamine	Undetermined	1	P	250.0000	
Chloroplatinic acid	Undetermined		P	250.0000	
Sodium bromate	Undetermined	1	P	500.0000	
Histamine	TTC	3	P	750.0000	
Isoniazid	NTTC	1	N		
Triethanolamine	Undetermined		N		
Magnesium sulfate	Undetermined		N		
Rapamycin	TTC	1, 3	N		
Mizoribine	Undetermined		N		
Warfarin	TTC	3	N		
2,4-Diaminotoluene	NTTC		N		
Cyclophosphamide	TTC	1	N*		
Dibenzopyrene	Undetermined		N		
Ethanol	TTC	1, 3	N		
Hexachlorobenzene	Undetermined		N		
Lithium carbonate	TTC	1,3	N		
Methanol	NTTC		N		
Methotrexate	TTC	3	N		
Dimethyl sulfoxide	NTTC		N		
Trichloroethylene	NTTC		N		
Mycophenolic acid	Undetermined		P		0.395061728
2-Mercaptobenzothiazole	Undetermined		P		16.11328125
Ribavirin	TTC	1, 3	P		26.04166667
Nicotinamide	Undetermined		P		288.0658436
Acetaminophen	Undetermined		P		288.0658436

P : Positive, N : No effect,

Blue color: accurate, Red color: false, yellow color: Undetermined because of insufficient reported data.

#: The criterion number used to define immunotoxicity

*: cyclophosphamide needs metabolic activity to demonstrate the activity

10-5. Factors responsible for false negative results in the IL-2 Luc assay

Although the within- and between-laboratory reproducibility satisfied the acceptance criteria for the validation study, the predictivity was less than 80%. We considered at least 2 reasons for the poor predictivity of the assay.

1) We collected immunotoxic information on the chemicals as much as possible and determined whether the chemicals exhibited T-cell dependent immunotoxicity or not using the criteria we proposed. The information used for classification were the effects of the chemicals on thymus weight, the production of cytokines predominantly produced by T cells, *in vitro* or *ex vivo*, T cell proliferation, and their reported mode of action on T cell function. However, the information available was very limited for most chemicals and very little data had been reproduced by different laboratories. The classification of some chemicals may not be correct.
2) The IL-2 Luc assay does not cover every aspect of the effects of the chemicals on T cell function. Other assays targeting T cell functions may be mandatory.

10-6. The applicability domain and the imitations of the IL-2 Luc assay

The IL-2 Luc assay evaluates the effects of chemicals on IL-2 transcription by T cells. Therefore, its applicability domain is immunotoxic chemicals the toxicity of which is caused by the direct effects of chemicals on T cells.

On the other hand, since the 2H4 cell line used in the IL-2 Luc assay is derived from Jurkat cells, a human acute T lymphoblastic leukemia cell line, it is conceivable that this cell line is more resistant to the cytotoxic effects of chemicals than bone marrow cells.

Therefore, the IL-2 Luc assay cannot evaluate the immunotoxic effects of some immunosuppressive drugs the mechanism of which is inhibiting DNA synthesis leading to myelotoxicity (Kimura et al., 2014). Thus, these chemicals in addition to chemicals that need metabolic activation should be outside the applicability domain. To overcome this drawback at present, the IL-2 Luc assay must be combined with assays capable of detecting myelotoxicity, such as *in vitro* myelotoxicity tests (Pessina et al., 2003). Similar to other *in vitro* test methods, poor water soluble chemicals are not suitable for this assay.

10-7. Potential of the IL-2 Luc assay

The IL-2 Luc assay evaluates the effects of chemicals on IL-2 transcription by Jurkat T cells stimulated with PMA and CI. The simultaneous stimulation of PMA and calcium ionophore or ionomycin surrogates the stimulation by T cell receptor (TCR) and CD28 (Kumagai et al., 1987; Truneh et al., 1985). The downstream signaling after the stimulation by TCR/CD28 is shown in Fig. 16. It indicates that the signaling required for IL-2 transcription after TCR/CD28 or PMA/CI stimulation involves the pathways leading the activation of AP1/2, mTOR, NF-κB, and NFAT. The immune system is composed of innate immune system and acquired immune system at least. The innate immune systems are activated by pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patters via Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), Nod-like receptors (NLRs), or cytokine receptors for IL-1 family or TNF family. Most of the downstream signaling after the stimulation of these receptors involves NF-κB and AP1/2 pathways (Newton and Dixit, 2012). In the acquired immune system, in addition to the process of T cell activation, B cell activation after B cell receptor stimulation and the signaling of various cytokines also involves NF-κB pathway (reviewed by Zhang and Sun (Zhang and Sun, 2015). Therefore, it is conceivable that the effects of chemicals on quite a few aspects of immune responses can be detected by the IL-2 Luc assay.

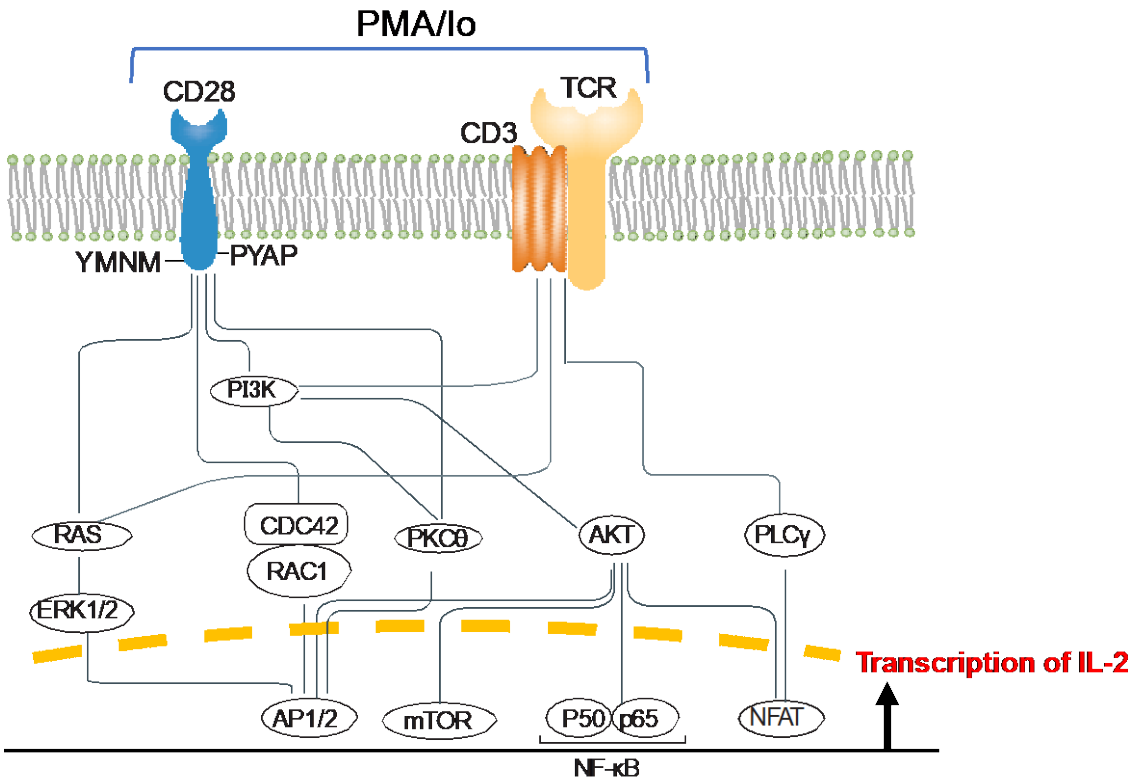


Fig. 16. The schematic presentation of cellular signaling after TCR/CD28 or PMA/Io stimulation.

Luster et al (Luster, 1988) proposed a screening battery using a 'tier' approach for detecting potential immunotoxic compounds in mice. Then, they defined criteria to classify immunotoxic chemicals using several parameters comprising the ‘tier approach’ and then, classified 51 chemicals into immunotoxic compounds or not (Luster et al., 1992b). Furthermore, they examined the ability of various immune tests to predict increased susceptibility in the host resistance classification (Luster et al., 1992a). Their final results demonstrated the following. 1. a number of the immune tests provided a relatively high association with changes in host resistance (i.e. > 70%) such as IgM plaque forming cell (PFC) response to sheep red blood cells, T cell mitogen response,

delayed hypersensitivity response (DHR), surface markers and spleen cellularity while several of the tests, such as leukocyte counts and lymphoproliferative response to LPS, were poor predictors with concordance values of approximately 50%. 2. The combinations of two immune tests compared with the host resistance classification increased the concordance from that obtained using individual tests. Pair-wise combinations which included either the PFC response, surface markers or DHRs gave consistently higher concordances.

When the IL-2 Luc assay examined 31 of the 51 chemicals evaluated by Luster et al. (1992b), its performance was similar to that of mixed lymphocyte reaction (MLR), DHR, and spleen cellularity and better than leukocyte counts or LPS response. Moreover, among 7 chemicals judged as false negative by the IL-2 Luc assay, 5 chemicals was judged as positive by Luster et al. (1992b) based on their suppressive effects on T cell mitogen response. Since our previous study demonstrated the inability of the IL-2 Luc assay to detect immunosuppressive effects of chemicals which are dependent on their suppressive effects on T cell proliferation, these 5 chemicals are out of applicability domain. Taking this into account, the sensitivity, specificity and accuracy of the IL-2 Luc assay was 76.5% (13/17), 44.4% (4/9), and 65.4% (17/26).

The HWBCRA, previously used in a rigorous prevalidation effort by ECVAM and other groups, is an immune test to examine the effects of chemicals on IL-4 or IL-1β production stimulated by staphylococcal enterotoxin B (SEB) or LPS, respectively (Langezaal et al., 2002). Although this study uses human whole blood cells, it examines the production of IL-4 by T cells and of IL-1 by monocytes. This concept is similar to that of the MITA, in which the effects of chemicals on T cells and monocytes are examined using Jurkat cell-derived 2H4 and THP-1-derived THP-G1b cells. Interestingly, the evaluation of chemicals by IL-4 production in the HWBCRA was almost identical to the results of the IL-2 Luc assay: both detected strong immunosuppression by FK506, cyclosporin A, dexamethasone and actinomycin D, which are more potent that chloroquine and azathioprine. Cyclophosphamide and

mizoribine require metabolic activation and thus are not considered as immunosuppressive by both assays. On the other hand, the cardiac glycoside digoxin is classified as an immunotoxic chemical by both assays. These data suggest that the IL-2 Luc assay may be an alternative method to the HWBCRA for examining the effects of chemicals on T cells. In addition, the IL-2 Luc assay has the following advantages over the HWBCRA. 1) The IL-2 Luc assay does not require primary cells, 2) it does not require cytokine quantification using ELISA, and 3) the time required for the IL-2 Luc assay is less than 8 h.

Finally, The performance of the IL-2 Luc assay to examine only immunosuppressive drugs whose effects on human are well established (reviewed by Allison (Allison, 2000)) showed that tacrolimus (TAC), cyclosporine A (CyA) and dexamethasone (Dex) significantly suppressed IL-2 luciferase activity (IL-2LA), although the average Lowest Observed Effect Levels (LOELs) of TAC and CyA were significantly lower that of DEX. The off-label immunosuppressive drugs chloroquine, minocycline and dapsone significantly suppressed IL-2LA. The anti-cancer drugs actinomycin D and cisplatin also significantly suppressed IL-2LA. In addition, azathioprine and colchicine were demonstrated to suppress IL-2LA. No suppressive effects on IL-2LA were demonstrated by several immunosuppressants which inhibit DNA synthesis or anti-proliferative effects on T cells, such as rapamycin, mizoribine, cyclophosphamide, methotrexate and mycophenolic acid.

10-8. Evaluation of the immunotoxicity of 60 chemicals by the modified MITA (mMITA)

Regulatory authorities worldwide require testing for allergic contact dermatitis (ACD) and appropriate hazard labeling to minimize exposures. Thus, we combined the MITA with an *in vitro* sensitization test, the IL-8 Luc assay, recently approved as an OECD test guideline for *in vitro* skin sensitization testing (OECD TG442E)(OECD, 2017). We designated this combined assay ‘modified MITA’ (mMITA). We established a

data set of 60 chemicals by referring to the publication by Wagner et al. (Wagner et al., 2006) in which they examined 46 chemicals characterized to different degrees for their immunotoxic and immunomodulatory properties using the Fluorescent Cell Chip (FCP) assay. In addition, we also evaluated the chemicals listed in the case studies in the Guidance for Immunotoxicity Risk Assessment for Chemicals published by World Health Organization (WHO)/ and Meeting, 2012. Since there were several overlaps between the chemicals we examined in our previous publication and those examined by the FCP, our final data set comprised 60 chemicals evaluated by the mMITA (Kimura et al., 2018) (Table 24). Table 25 lists the chemicals that affected the normalized IL-2 luciferase activity in increasing order of their Lowest Observed Effect Level (LOEL), the results of the MITA evaluation (suppression (S), augmentation (A) or no effect (N)), the LOEL for each parameter of each chemical, and the results of the IL-8 Luc assay evaluation (sensitiser (S) and non-sensitiser (N)).

Table 24. Classification of chemicals by the mMITA in increasing order of the LOEL of the IL-2 Luc assay.

Chemicals	IL-2		IFN-γ		IL-1β		IL-8		IL-8 Luc
	Judge	LOEL	Judge	LOEL	Judge	LOEL	Judge	LOEL	Judge
FK 506	S	0.00	S	0.00	A		N		N
Cyclosporine A	S	0.00	S	0.00	N		N		N
Actinomycin D	S	0.00	S	0.01	N		S	0.00	S
Digoxin	S	0.01	S	0.02	N		N		S
Dexamethasone	S	0.01	N		S	0.01	S	0.01	N
Dibenzopyrene	S	0.01	S	0.03	N		N	0.00	N
Pyrimethamine	S	0.04	N		N		N		N
Chloroquine	S	0.05	S	0.02	S	10.00	S	30.00	S
Cisplatin	S	0.24	S	1.22	N		N		S
Hydrocortisone	S	0.34	A	6.27	S	0.34	S	0.34	N
Mitomycin C	S	0.36	N		N		N		S
Citral	S	0.36	S	1.37	N		N		S
Nitrofurazone	S	0.37	A	3.91	A		A	62.50	S
FR167653	S	0.49	S	0.49	S	145.83	S	125.00	N
Amphoterycin B	S	0.78	S	2.08	A	3.13	A	7.82	S
2-Aminoanthracene	S	0.81	S	5.86	S	2.03	N		S
Lithium carbonate	S	0.98	A	116.67	S	0.39	S	0.39	S
Isophorone diisocyanate	S	0.98	N		S	0.98	S	0.98	S
p-Nitroaniline	S	0.98	S	1.95	S	1.47	S	2.45	N
Dibutyl phthalate	S	0.98	S	1.95	S	39.07	S	31.25	N
Formaldehyde	S	1.71	N		S	15.63	S	15.63	S
Benzethonium chloride	S	1.95	S	1.95	S	3.91	N		S
Isoniazid	S	1.97	N		N		S	800.00	N
Chlorpromazine	S	3.91	S	3.91	S	7.81	S	7.81	S
Cobalt chloride	S	3.91	S	9.12	S	3.91	S	125.00	S
Pentamidine isethionate	S	3.91	S	32.55	S	3.91	S	3.91	N
Aluminum chloride	S	3.91	S	62.50	N		N		N
Lead(II) acetate	S	3.91	S	3.91	N		N		N
Hydrogen peroxide	S	7.82	S	31.25	N		N		S
Minocycline	S	8.33	S	5.00	N		N		S
Histamine	S	9.12	A	5.86	N		S	3.91	S
Diethanolamin	S	9.12	N		N		N		S
Nickel sulfate	S	14.32	S	32.55	S	250.00	S	250.00	S
Sulfasalazine	S	36.00	S	1.20	S	7.80	S	1.20	N
Diesel exhaust particles	S	39.07	A	47.53	N		S	62.50	S
Dapsone	S	45.01	S	55.14	S	46.88	S	134.75	N
Sodium bromate	S	125.00	N		N		N		S
Triethanolamine	S	187.50	S	1416.67	N		N		S
Mercuric chloride	**N**		**A**	**3.91**	**S**	**1.95**	**S**	**1.95**	**S**
Chloroplatinic acid	**N**		**N**		**N**		**S**	**15.63**	**S**
2-Mercaptobenzothiazole	**N**		**N**		**N**		**S**	**125.00**	**S**
Cyclophosphamide	**N**		**A**	**168.00**	**N**		**N**		**S**
Magnesium sulfate	**N**		**N**		**S**	**15.63**	**N**		**S**
Sodium dodecyl sulfate	**N**		**N**		**N**		**N**		**S**
2,4-Diaminotoluene	**N**		**A**	**62.50**	**N**		**S**	**0.98**	**N**
Ethanol	**N**		**N**		**N**		**N**		**N**
Methanol	**N**		**N**		**N**		**N**		**N**
Hexachlorobenzene	**N**		**N**		**N**		**N**		**N**
Trichloroethylene	**N**		**N**		**N**		**N**		**N**
Azathioprine	**N**		**A**	**40.01**	**A**	**9.23**	**N**		**N**
Mizoribine	**N**		**N**		**A**	**5.20**	**A**	**7.45**	**N**
Rapamycin	A	0.00	N		A	0.91	N		S
Nicotinamide	A	0.10	A	110.03	S	3.00	S	10.00	N
Colchicine	A	0.29	A	0.06	A	0.02	A	20.00	S
Mycophenolic acid	A	0.38	A	6.24	N		N		S
Methotrexate	A	0.45	A	0.09	N		N		N
Dimethyl sulfoxide	A	3.91	A	625.00	S	66.41	S	3.91	N
Ribavirin	A	15.63	A	187.50	A	5.86	N		N
Warfarin	A	23.33	N		S	30.00	S	0.00	N
Acetaminophen	A	33.33	A	33.33	A	166.67	A	100.00	N

Table 25. The group by LOEL

Groups	Suppression of IL-2 promoter activity (LOEL ∝g/ml)
Group 1	LOEL<0.1
Group 2	0.1≤LOEL<1.0
Group 3	1.0≤LOEL<10
Group 4	10≤LOEL<1000
Group 5	None
Group 6	Augmentation

0.0 of the LOEL means less than 0.001.

Using this data set, we first demonstrated a significant correlation between LOELs for the effects on the IL-2 luciferase assay and those on the IFN luciferase assay, and between LOELs for effects on the IL-1β luciferase assay and those on the IL-8 luciferase assay (Kimura et al., 2018) (Fig. 17). These results indicated that evaluations of the effects of chemicals on the IL-2 and IL-8 luciferase assays can provide immunotoxicological information almost equivalent to the evaluation of these chemicals using the IL-2, IFN-γ, IL-1β, and IL-8 luciferase assays.

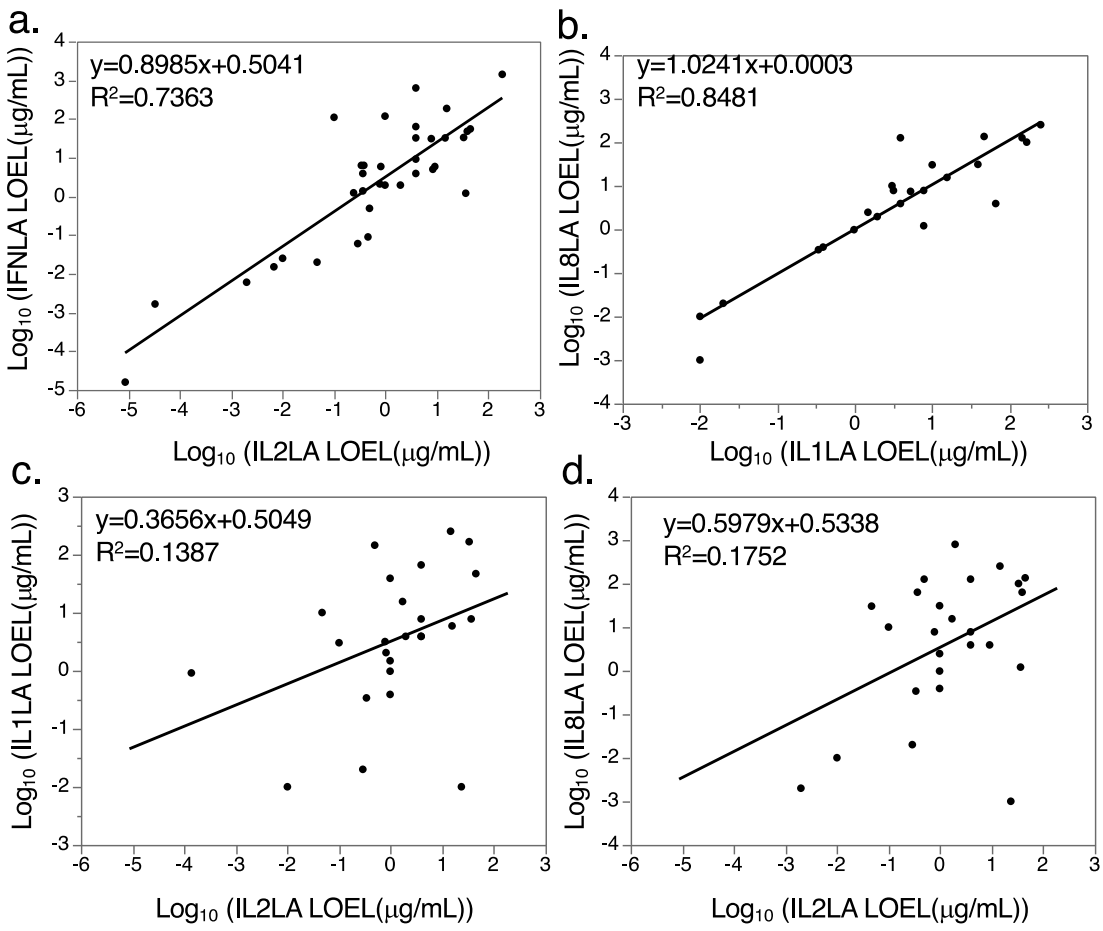
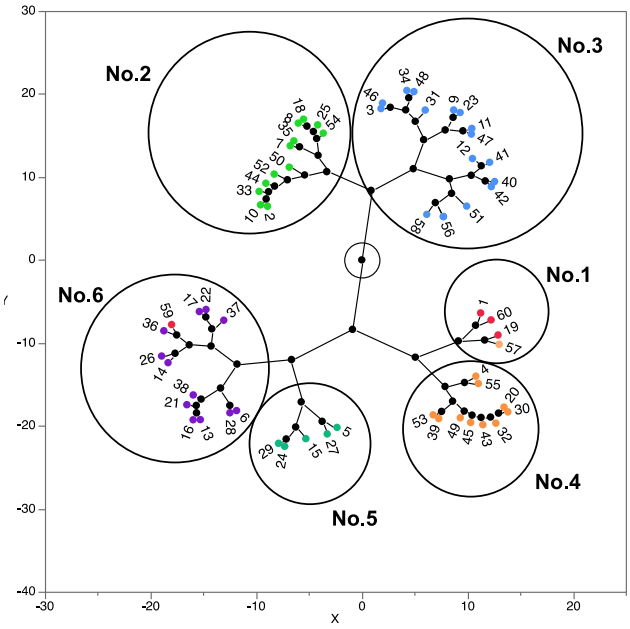


Fig. 17. The correlation between the LOEL for the 4 luciferase assays.

Next, we demonstrated that K-means clustering and hierarchical clustering of the 60 chemicals based on the LOEL for their effects on IL-2 and IL-8 promoter activities, and the judgment by the IL-8 Luc assay, resulted in the same 6-cluster solution: cluster 1 with preferential suppression of IL-8, cluster 2 with suppression of IL-2 and a positive IL-8 Luc assay result, cluster 3 with suppression of both IL-2 and IL-8, cluster 4 with no effects on IL-2 or IL-8 and a negative IL-8 Luc assay result, cluster 5 with suppression of both IL-2 and IL-8 and a negative IL-8 Luc assay result, and cluster 6 with preferential suppression of IL-2 (Kimura et al., 2018) (Figs. 18, 19 and 20). These data suggest that the mMITA is a promising novel high-throughput approach for detecting unrecognized immunological effects of chemicals and for profiling their

immunotoxic effects. The data obtained from these assays can be used by both industry and regulatory agencies to assess the immunotoxicity risks of chemicals. Toward this particular goal, the IL-2 Luc assay and the IL-8 or IL-1β Luc assay should be officially validated and a larger number of chemicals must be evaluated using the MITA to fully determine the potential and limits of this technique.



Cluster No	No	Chemicals
1	1	2,4-Diaminotoluene
1	19	Dimethyl sulfoxide
1	57	Nicotinamide
1	60	Warfarin
2	2	2-Aminoanthracene
2	7	Amphoterycin B
2	8	Benzethonium chloride
2	10	Cisplatin
2	18	Diethanolamin
2	25	Hydrogen peroxide
2	33	Mitomycin C
2	35	Nitrofurazone
2	44	Citral
2	50	Chloroquine
2	52	Digoxin
2	54	Minocycline
3	3	2-Mercaptobenzothiazole
3	9	Chlorpromazine
3	11	Cobalt chloride
3	12	Cyclophosphamide
3	23	Histamine
3	31	Mercuric chloride
3	34	Nickel sulfate
3	40	Sodium bromate
3	41	Sodium dodecyl sulfate
3	42	Triethanolamine
3	46	Chloroplatinic acid
3	47	Formaldehyde
3	48	Diesel exhaust particles
3	51	Colchicine
3	56	Mycophenolic acid
3	58	Rapamycin

Cluster No	No	Chemicals
4	4	Acetaminophen
4	20	Ethanol
4	30	Magnesium sulfate
4	32	Methanol
4	39	Ribavirin
4	43	Hexachlorobenzene
4	45	Trichloroethylene
4	49	Azathioprine
4	53	Methotrexate
4	55	Mizoribine
5	5	Actinomycin D
5	15	Dexamethasone
5	24	Hydrocortisone
5	27	Isophorone diisocyanate
5	29	Lithium carbonate
6	6	Aluminum chloride
6	13	Cyclosporine A
6	14	Dapsone
6	16	Dibenzopyrene
6	17	Dibutyl phthalate
6	21	FK 506
6	22	FR167653
6	26	Isoniazid
6	28	Lead(II) acetate
6	36	Pentamidine isethionate
6	37	p-Nitroaniline
6	38	Pyrimethamine
6	59	Sulfasalazine

Fig. 18. Hierarchical clustering of 60 chemicals by the mMITA

Hierarchical clustering of 60 chemicals was performed for these 3 immunotoxic parameters and visualized using JMP pro 13.1.0. Table is the list of chemicals that belong to each cluster.

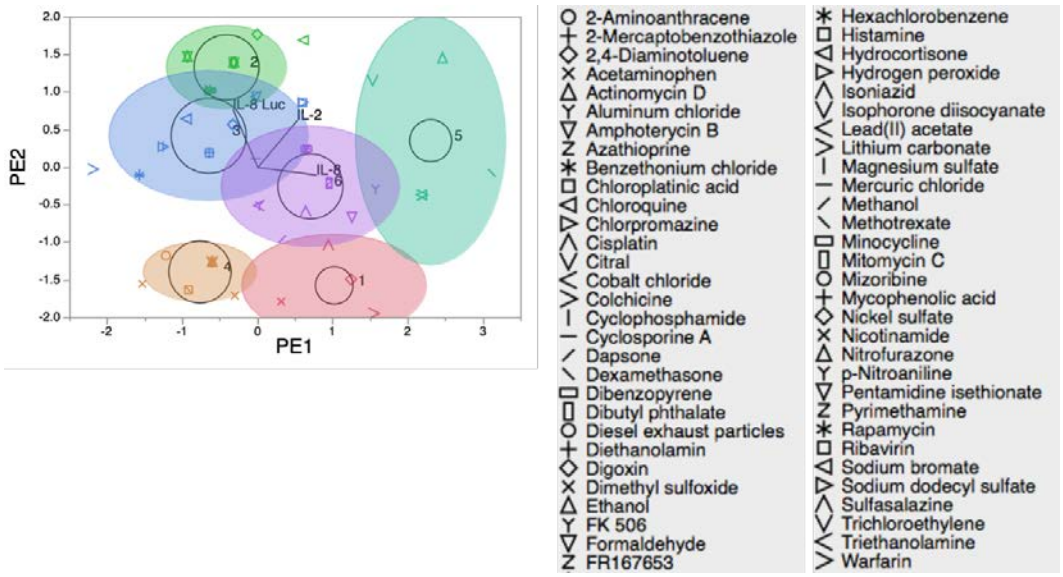


Fig. 19. K-means clustering analysis of chemicals by MITA

K-means clustering of 60chemicals was performed for these 3 immunotoxic parameters and visualized using JMP pro 13.1.0.

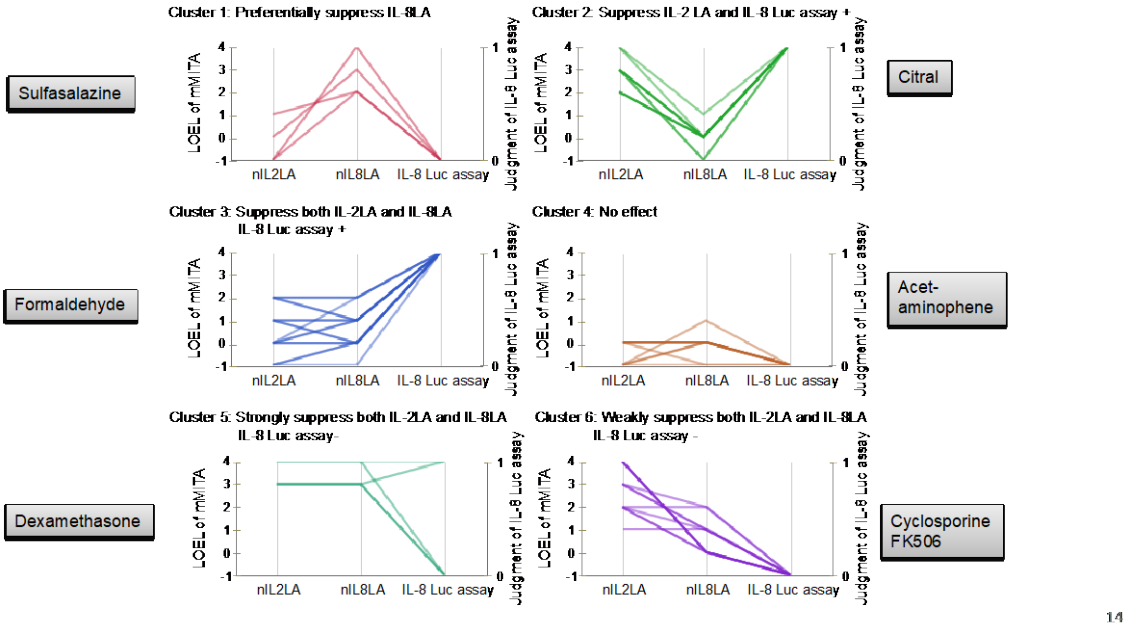


Fig. 20. Characteristics of each cluster and their representative chemicals

The scores for the LOEL of IL2LA, IL8LA and the IL8 Luc assay was plotted for each chemical belonging to different clusters.

10-9. The regulatory application of the IL-2 Luc assay.

The CAS REGISTRY^{SM} currently contains more than 130 million unique organic and inorganic chemical substances, such as alloys, coordination compounds, minerals, mixtures, polymers, and salts. Humans are exposed to many of these substances, which are present as environmental contaminants or used as food additives and drugs. Some of these compounds can target the immune system, resulting in adverse health effects such as the development of allergies, autoimmune disorders, increased susceptibility to infection and cancer, and other diseases. Accordingly, immunotoxicity, which is defined as the toxicological effects of xenobiotics on the function of the immune system, is a matter of serious concern to the public as well as regulatory agencies. To address these concerns, the World Health Organization published its Guidance for Immunotoxicity Risk Assessment for Chemicals (WHO). Currently, the assessment of chemical immunotoxicity relies mainly on animal models and assays that characterize immunosuppression and sensitization. However, animal studies have so many drawbacks, such as high cost, ethical concerns, and questionable relevance to risk assessment for humans, that they cannot screen immunotoxicity of more than 130 million chemicals. Therefore, it is an urgent matter to develop alternative testing methods and assessment strategies to reduce the use of laboratory animals and, if possible, replace animals used in scientific studies (Adler et al., 2011). So far, however, there is no OECD test guidelines to detect chemical immunotoxicity in vitro. Therefore, we would like to propose the IL-2 Luc assay, and the MITA in near future, as a screening toolbox of alternative test methods for immunotoxicity.

Finally, the VMT recommend that the proficiency chemicals (Appendix 15) to users and the performance standard chemicals (Appendix 16) to me-too validation study.

11. Conclusion

In this study, we conducted the validation study of the IL-2 Luc assay among the 4 luciferase assays that comprise the MITA. The results of both Phase I and Phase II studies satisfied the acceptance criteria for the validation study. Although the predictivity could not reach 80%, it may be acceptable when considering its applicability domain and limited target. So, we would like to propose the IL-2 Luc assay for the OECD test guideline of *in vitro* immunotoxicity test.

12. Acknowledgement

This validation study was supported by the Grants-in-Aid for the Ministry of Economy, Trade and Industry (METI), the Ministry of Health, Labor and Welfare (MHLW) and the Japanese Society for Alternatives to Animal Experiments (JSAAE). We gratefully acknowledge the voluntary works by the participated laboratories and the European Union Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM), and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

13. References

Adler, S., Basketter, D., Creton, S., et al. (2011), Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. *Arch Toxicol* 85: 367-485, 10.1007/s00204-011-0693-2

Alegre, M.L., Frauwirth, K.A., Thompson, C.B. (2001), T-cell regulation by CD28 and CTLA-4. *Nat Rev Immunol* 1: 220-228, 10.1038/35105024

Allison, A.C. (2000), Immunosuppressive drugs: the first 50 years and a glance forward. *Immunopharmacol* 47: 63-83,

Corsini, E., Roggen, E. (2017), Overview of in vitro assessment of immunotoxicity. *DOI:10.1016/j.cotox.2017.06.016*

Elentner, A., Schmuth, M., Yannoutsos, N., et al. (2018), Epidermal Overexpression of Xenobiotic Receptor PXR Impairs the Epidermal Barrier and Triggers Th2 Immune Response. *J Invest Dermatol* 138: 109-120, 10.1016/j.jid.2017.07.846

Fulda, S., Gorman, A.M., Hori, O., et al. (2010), Cellular stress responses: cell survival and cell death. *Int J Cell Biol* 2010: 214074, 10.1155/2010/214074

Galbiati, V., Mitjans, M., Corsini, E. (2010), Present and future of in vitro immunotoxicology in drug development. *J Immunotoxicol* 7: 255-267, 10.3109/1547691X.2010.509848

Gallucci, S., Matzinger, P. (2001), Danger signals: SOS to the immune system. *Curr Opin Immunol.* 13: 114-119,

Gennari, A., Ban, M., Braun, A., et al. (2005), The Use of In Vitro Systems for Evaluating Immunotoxicity: The Report and Recommendations of an ECVAM Workshop. *J Immunotoxicol* 2: 61-83, 10.1080/15476910590965832

Hidaka, T., Ogawa, E., Kobayashi, E.H., et al. (2017), The aryl hydrocarbon receptor AhR links atopic dermatitis and air pollution via induction of the neurotrophic factor artemin. *Nat Immunol* 18: 64-73, 10.1038/ni.3614

Kaiko, G.E., Horvat, J.C., Beagley, K.W., et al. (2008), Immunological decision-making: how does the immune system decide to mount a helper T-cell response? *Immunology* 123: 326-338, 10.1111/j.1365-2567.2007.02719.x

Kimura, Y., Fujimura, C., Ito, Y., et al. (2014), Evaluation of the Multi-ImmunoTox Assay composed of 3 human cytokine reporter cells by examining immunological effects of drugs. *Toxicol In Vitro* 28: 759-768, 10.1016/j.tiv.2014.02.013

Kimura, Y., Fujimura, C., Ito, Y., et al. (2018), Profiling the immunotoxicity of chemicals based on in vitro evaluation by a combination of the Multi-ImmunoTox assay and the IL-8 Luc

assay. *Arch Toxicol* 92: 2043-2054, 10.1007/s00204-018-2199-7

Kimura, Y, Yasuno R, Watanabe M., et al. (2020), An international validation study of the IL-2 Luc assay for evaluating the potential immunnotoxic effects of chemicals on T cells and a proposal for reference data for immunotoxic chemicals. Toxicol in Vitro in press, 10.1016/j.tiv.2020.104832

Kultz, D. (2005), Molecular and evolutionary basis of the cellular stress response. *Ann Rev Physiol* 67: 225-257, 10.1146/annurev.physiol.67.040403.103635

Kumagai, N., Benedict, S.H., Mills, G.B., et al. (1987), Requirements for the simultaneous presence of phorbol esters and calcium ionophores in the expression of human T lymphocyte proliferation-related genes. *J Immunol* 139: 1393-1399,

Langezaal, I., Hoffmann, S., Hartung, T., et al. (2002), Evaluation and prevalidation of an immunotoxicity test based on human whole-blood cytokine release. *ATLA* 30: 581-595,

Lankveld, D.P., Van Loveren, H., Baken, K.A., et al. (2010), In vitro testing for direct immunotoxicity: state of the art. *Meth Mol Biol* 598: 401-423, 10.1007/978-1-60761-401-2_26

Liao, W., Lin, J.X., Wang, L., et al. (2011), Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. *Nat Immunol* 12: 551-559, 10.1038/ni.2030

Luster, M.I., Munsosn, A.E., Thomas, P.D., et al. (1988), Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. *Fundam Appl Toxicol* 10: 2-19, 10.1016/0272-0590(88)90247-3

Luster, M.I., Portier, C., Pait, D.G., et al. (1992a), Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. *Fundam Appl Toxicol* 18: 200-210, 10.1016/0272-0590(92)90047-l

Luster, M.I., Pait, D.G., Portier, C., et al. (1992b), Qualitative and quantitative experimental models to aid in risk assessment for immunotoxicology. *Toxicol Lett* 64-65 Spec No: 71-8, 10.1016/0378-4274(92)90174-i

Luster, M.I., Portier, C., Pait, D.G., et al. (1993), Risk assessment in immunotoxicology. II. Relationships between immune and host resistance tests. *Fundam Appl Toxicol* 21: 71-82,

Michelini, E., Cevenini, L., Calabretta, M.M., et al., (2014). Exploiting in vitro and in vivo bioluminescence for the implementation of the three Rs principle (replacement, reduction, and refinement) in drug discovery. Anal Bioanal Chem 406, 5531-5539, 10.1007/s00216-

014-7925-2

Milisav, I., (2011), Cellular stress responses. Advance in regererative medicine ect. Sabine Wislet-Gendebien InTech, 978-953-307-732-1

Nakajima, Y., Ikeda, M., Kimura, T., et al. (2004), Bidirectional role of orphan nuclear receptor RORalpha in clock gene transcriptions demonstrated by a novel reporter assay system. *FEBS Lett* 565: 122-126, 10.1016/j.febslet.2004.03.083

Nakajima, Y., Kimura, T., Sugata, K., et al. (2005), Multicolor luciferase assay system: one-step monitoring of multiple gene expressions with a single substrate. *Biotechniques* 38: 891-894, 05386ST03 [pii]

Nakajima Y, Ohmiya Y., (2010), Bioluminescence assays: multicolor luciferase assay, secreted luciferase assay and imaging luciferase assay. *Expert Opin Drug Discov* 5:835-49, 10.1517/17460441.2010.506213

Newton, K., Dixit, V.M. (2012), Signaling in innate immunity and inflammation. *Cold Spring Harb Perspect Biol* 4, 10.1101/cshperspect.a006049

Niwa, K., Ichino, Y., Kumata, S., et al. (2010), Quantum yields and kinetics of the firefly bioluminescence reaction of beetle luciferases. *Photochem Photobiol* 86: 1046-1049, 10.1111/j.1751-1097.2010.00777.x

Noguchi, T., Ikeda, M., Ohmiya, Y., et al. (2008), Simultaneous monitoring of independent gene expression patterns in two types of cocultured fibroblasts with different color-emitting luciferases. *BMC Biotechnol* 8: 40, 1472-6750-8-40 [pii]10.1186/1472-6750-8-40

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 (2017), OECD Test Gudeline for the Testing of Chemicals No.442E: In Vitro Skin Sensitisation assays addressing the key Event on activation of dendritic cells on the Adverse Outcome pathway for Skin Sensitisation. http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

Pessina, A., Albella, B., Bayo, M., et al. (2003), Application of the CFU-GM assay to predict acute drug-induced neutropenia: an international blind trial to validate a prediction model for the maximum tolerated dose (MTD) of myelosuppressive xenobiotics. *Toxicol Sci* 75: 355-367, 10.1093/toxsci/kfg188

Roda, A., Pasini, P., Mirasoli, M., et al. (2004), Biotechnological applications of bioluminescence and chemiluminescence. *Trends Biotechnol* 22: 295-303, 10.1016/j.tibtech.2004.03.011

Saito, R., Hirakawa, S., Ohara, H., et al. (2011), Nickel differentially regulates NFAT and NF-kappaB activation in T cell signaling. *Toxicol Appl Pharmacol* 254: 245-255, 10.1016/j.taap.2011.04.017

Sasaki, Y., Aiba, S. (2007), Dendritic cells and contact dermatitis. *Clinl Rev Allergy Immunol* 33: 27-34, 10.1007/s12016-007-0034-7 [doi]

Takahashi, T., Kimura, Y., Saito, R., et al. (2011), An in vitro test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. *Toxicol Sci* 124: 359-369, 10.1093/toxsci/kfr237

Truneh, A., Albert, F., Golstein, P., et al. (1985), Calcium ionophore plus phorbol ester can substitute for antigen in the induction of cytolytic T lymphocytes from specifically primed precursors. *J Immunol* 135: 2262-2267,

Wagner, W., Walczak-Drzewiecka, A., Slusarczyk, A., et al. (2006), Fluorescent Cell Chip a new in vitro approach for immunotoxicity screening. *Toxicol letters* 162: 55-70, 10.1016/j.toxlet.2005.10.017

W.H.O., Meeting, I.P.o.C.S.I.S., (2012), Guidance for immunotoxicity risk assessment for chemicals. WHO press,

Zhang, H., Sun, S.C. (2015), NF-kappaB in inflammation and renal diseases. *Cell Biosci* 5: 63, 10.1186/s13578-015-0056-4

14. List of abbreviations

95% CI : the 95% confidence interval

AIST : National Institute of Advanced Industrial Science and Technology

AOP : Adverse outcome pathway

ARE: Antioxidant response element

CAS No. : Chemical Abstract Service Number

CMV : Cytomegalovirus

CSC : the Chemical Selection Committee

DMSO : Dimethyl sulphoxide

DPRA : the Direct Peptide Reactivity Assay

ECVAM : the European Centre for Validation of Alternative Methods

EDTA : Ethylenediaminetetraacetic acid

EGFR : Epidermal growth factor receptor

EGR-1 : Early growth response-1

EU : European Union

FBS : Fetal bovine serum

FN : False Negative Rate

GLP : Good laboratory Practice

GSH : Glutathione

HRI/FDSC : Hatano Research Institute, Food & Drug Safety Center

HSV : Herpes simplex viruses

ICCVAM : Interagency Coordinating Committee on the Validation of Alternative Methods

ID : Identification

IFN-γ : Interferon-γ

Inh-GAPLA : Inhibition index of GAPLA

IL-2 : Interleukin-2

IL-8 : Interleukin-8

JaCVAM : the Japanese Center for the Validation of Alternative Methods

Keap-1 : Kelch-like ECH-associated protein 1

KoCVAM : Korean Center for the Validation of Alternative Methods

LLNA : Local lymph node assay

LPS : Lipopolysaccharide

MIT : Minimum induction threshold

MITA : Multi-Immuno Tox Assay

mMUSST : modified myeloid U937 dendritic cell activation test

MoDCs : Monocyte-derived dendritic cells

MOVS: Management Office of Validation Study

mRNA : messenger ribonucleic acid

MSDS : Material safety data sheet

NICEATM : the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

NIHS : National Institute of Health Sciences

NPV : Negative predictive value

Nqo1 : NADPH-quinone oxidoreductase 1

Nrf2 : Nuclear factor (erythroid-derived 2)-like factor 2

nIL2LA : normalized IL2LA

nIFNLA : normalized IFNLA

OECD : the Organization for Economic Co-operation and Development

PCR : Polymerase chain reaction

PI : Propidium iodide

PMA/Io : Phorbol 12-myristate 13-acetate/Ionomycin

PN : False Positive Rate

PPV : Positive Predictive Value

QC : Quality Control

REACH : Registration, Evaluation, Authorization and Restriction of CHemicals

RFI : Relative fluorescence intensity

RT : Ring trial

SLG : Stable luciferase green

IL2LA : SLG luciferase activity

SLO : Stable luciferase orange

IFNLA : SLO luciferase activity

SLR : Stable luciferase red

GAPLA : SLR luciferase activity

SLS : Sodium lauryl sulfate

SLR : Stable luciferase red

SLR-LA : SLR luciferase activity

SV40 : Simian virus 40

TG : Test Guideline

TNF-α : Tumor necrosis factor-α

UN GHS : United Nations Globally Harmonized System of Classification and Labeling of Chemicals

VMT : Validation Management Team

15. Appendixes