Report on a Validation Study of the IL-2 Luc leukocyte toxicity test (IL-2 Luc LTT)

MITA Validation Management Team

1. Summary 6
3. Background 13
3-1. What is immunotoxicity? 13
3-2. The current status of *in vitro* immunotoxicity tests 14
3-3. Predictivity of *in vitro* immunotoxicity tests 15
3-4. Multi-ImmunoTox Assay (MITA) 16
3-5. The evaluation of immunotoxicity profiles of well-known immunosuppressive drugs by the MITA 18
3-7. The IL-2 Luc assay plays a principal role in detecting the immunosuppressive effects of chemicals in the MITA 20
3-8. The IL-2 Luc assay cannot detect immunosuppressive effects of drugs which are dependent on the inhibition of DNA synthesis or anti-proliferative effects 20
3-9. The IL-2 Luc LTT can detect antimitotic effects of chemicals. 21
3-10. The process of validation of the combined assays of the IL-2 Luc assay and the IL-2 Luc LTT 23
4. Test method and modification 24
4-1. IL-2 reporter cell line 2H4 24
4-2. Chemical treatment of 2H4 cells and measurement of luciferase activity 24
4-3. Criteria to determine antimitotic effects of chemicals 26
4-4. Bioluminescence system 28
5. Validation Management Structure 31
5-1. Validation Management Team (VMT) 31
5-2. Management office 32
5-3. Meetings held 32
6. Study Design 33
7. Test Chemicals 34
7-1. Basic rules for chemical selection 34
7-1-1. Applied selection criteria 35
7-1-2. Chemical acquisition, coding and distribution 36
7-1-3. Handling 36
7-2. Pre-validation - Phase 0 trial 37
7-3. Validation study - Phase I trial 37
7-4. Validation study - Phase II trial 38

8. Protocols 40
8-1. Overview of IL-2 Luc LTT 40
8-2. Protocol for IL-2 Luc LTT 41
8-2-1. Reagents and equipment 41
8-2-2. Culture media 43
8-2-3. Preparation of stimulant for 2H4 cells 44
8-2-4. Thawing of 2H4 cells 45
8-2-5. Maintenance of 2H4 cells 45
8-2-6. Preparation of cells for assay 46
8-2-7. Preparation of chemicals and cell treatment with chemicals 46
8-2-8. Dilution of chemicals 48
8-2-9. Measurements 48
8-2-10. Luminometer apparatus 49
8-2-11. Positive and negative controls 49
8-2-12. Calculation and definition of parameters for IL-2 Luc LTT 49
8-2-13. Acceptance criteria 50
8-2-14. Criteria 51
8-3. Data collection 52
8-3-1. Operating procedure 52
8-3-2. Chemicals 53
8-3-3. Data handling 53
8-3-4. Index from each experiment and decision criteria for judgment 54
8-3-5. Reliability 56
8-3-6. Predictivity 57
8-4. Quality assurance 58
9. Results 58
9-1. Final criteria for Phase I trial 59
9-1-1. Acceptance criteria 59
9-1-2. Criteria 60
9-1-3. Predictivity 61
9-2. Phase 0 trial (for technical transfer) 63
9-3. Phase I trial (for within- and between-laboratory reproducibility) 63
9-3-1. Test conditions 63

9-3-2. Within-laboratory variation assessments in Phase I trial 64
9-3-3. Between-laboratory variation assessments in Phase I trial 64
9-4. Phase II trial (for between-laboratory reproducibility and predictivity) 65
9-4-1. Test conditions 65
9-4-2. Between-laboratory variation assessments in Phase II trial 66
9-5. Quality assurance (Appendix 16) 67
9-6. Combined results of Phase I and II trials (for between- and within- laboratory reproducibility) 69
9-6-1. Test conditions 69
9-6-2. Between-laboratory variation assessments from Phase I and II trials (Table 17) 69
9-7. The predictivity of the IL-2 Luc LTT in the Phase I and Phase II trials 72
10. Discussion 75
10-1. Reliability 75
10-2. Between- and within-laboratory reproducibility 76
10-3. Predictivity 76
10-3-1. The predictivity of the IL-2 Luc LTT in the validation trials 76
10-3-2. IL-2 Luc LTT judgment for 83 data set chemicals 76
10-3-3. Application of the IL-2 Luc LTT with IL-2 Luc as part of an IATA approach 78
10-4. The possible reason for the false negative judgment for carmustin, paclitaxel, leflunomide, and methotrexate (MTX) by the IL-2 Luc LTT. 87
10-5. Limitation and applicability domain of the IL-2 Luc LTT 88
10-6. The possible reason for the poor specificity of the IL-2 Luc LTT 89
10-7. Regulatory application of the IL-2 Luc LTT 92
11. Conclusion 93
12. Acknowledgement 94
13. References 95
14. List of abbreviations 98

List of Appendixes.

Appendix 01 IL-2 Luc LTT Study plan Ver1.0
Appendix 02-1 Coded chemicals
Appendix 02-2 Coded chemicals
Appendix 03 Chemical structure of the test chemicals for the Phase 0 trial_IL-2 Luc LTT
Appendix 04 Chemical structure of the test chemicals for the Phase I trial_IL-2 Luc LTT
Appendix 05 Chemical structure of the test chemicals for the Phase II trial_IL-2 Luc LTT
Appendix 06 IL-2 Luc LTT protocol Ver.001.4
Appendix 07 Principle of measurement of luciferase activity
Appendix 08 Validation of reagents and equipment
Appendix 09 Data sheet for MITA Il-2 Luc LTT Ver.008.8 20210112
Appendix 10 MITA IL-2 Luc LTT Phase I results trial
Appendix 11 MITA IL-2 Luc LTT Phase II results trial
Appendix 12 Supplementary Table 1 IL-2 Luc LTT
Appendix 13 Reference 25 chemicals_ text
Appendix 14 Supplementary Table 2 IL-2 Luc LTT
Appendix 15 Reference 63 chemicals_ text
Appendix 16 QC for IL-2 Luc LTT validation (R1)
Appendix 17-1 IL-2 Luc LTT Check List
Appendix 17-2 IL-Luc LTT Check List (1)

1. Summary

The IL-2 Luciferase leukocyte toxicity test (IL-2 Luc LTT) was developed using one of the luciferase reporter assays that comprise the Multi-ImmunoTox Assay (MITA), a high-throughput screening system that our group developed to evaluate chemical immunotoxicity. The MITA assesses effects of chemicals on promoter activity of several cytokines (i.e., IL-1, IL-2, IFN-γ, and IL-8). Although the final long-term goal is to officially validate the MITA for within- and between-laboratory reproducibility and predictivity, so far the validation studies for the individual IL-2 luciferase assay (IL-2 Luc assay), the IL-1 luciferase assay (IL-1 Luc assay), and, as reported here, the IL-2 Luc LTT have been conducted step by step. This report describes the results of the validation study for the IL-2 Luc LTT.

In the MITA, we used three stable reporter cell lines transfected with luciferase genes under the control of IL-2 and IFN-β, IL-8, or IL-1β promoters. The IL-2 Luc LTT used 2H4 cells from among these cell lines, which were derived from Jurkat cells containing stable luciferase green (SLG) regulated by the IL-2 promoter, stable luciferase orange (SLO) regulated by the IFN-β promoter, and stable luciferase red (SLR) regulated by the glycerol-3-phosphate dehydrogenase (GAPDH) promoter.

The lead laboratory reported that the IL-2 Luc assay using 2H4 cells can detect the effect of chemicals on IL-2 transcription and hence the immunotoxicity of chemicals, mainly by affecting T cell function. In addition, the validation study for the IL-2 Luc assay conducted by three independent laboratories showed reasonable performance of

the assay. However, it was found that immunotoxicity due to antimitotic effects of the chemicals could not be detected by the IL-2 Luc assay.

To overcome this drawback of the IL-2 Luc assay, the lead laboratory attempted to establish a new assay to detect antimitotic effects of chemicals by taking advantage of the triple reporter cell properties of 2H4 cells. To begin with, it was hypothesized that if 2H4 cells are treated with antimitotic agents for 24 h and then stimulated with PMA (phorbol myristate acetate) /Io (ionomycin), they might reduce GAPDH promoter-driven luciferase activity (GAPLA) without suppressing IL-2 promoter-driven luciferase activity (IL2LA). We first determined the optimal incubation time with drugs and the seeding cell density, and then, the criteria to detect antimitotic effects of chemicals, namely the reduction of GAPLA without reducing IL2LA. We designed the assay as the IL-2 luciferase leukocyte toxicity test (IL-2 Luc LTT).

In the validation study of the IL-2 Luc LTT, the preliminary test trial (Phase 0) was performed by the participating laboratories following explicit explanations of "IL-2 Luc leukocyte toxicity test (IL-2 Luc LTT) protocol ver.001.2" by the lead laboratory, Tohoku University. Three laboratories participated in the Phase 0 trial of the IL-2 Luc LTT using three open-labeled chemicals, namely bleomycin sulfate, dexamethasone, and 6-thioguanine, and conducted one set of three experiments for each chemical. The response patterns for the three chemicals were similar among the three laboratories. Based on these results, the validation management team (VMT) judged the technical and protocol transfer of the IL-2 Luc LTT as acceptable.

In the Phase I trial, a total of five coded chemicals were evaluated by three experimental sets based on the IL-2 Luc LTT protocol ver.001.3. The within-laboratory reproducibility was 100.0% (15/15). The between-laboratory reproducibility was 100.0% (5/5).

In the Phase II trial, a total of 20 coded chemicals were evaluated by one experiment set based on the IL-2 Luc LTT protocol ver.001.4. The between-laboratory reproducibility was 90.0 % (18/20).

In the combined results of the Phase I and II trials, the average within-laboratory reproducibility was 100.0% (15/15). The between laboratory reproducibility was 92.0% (23/25).

The predictivity of the IL-2 Luc LTT was determined for detecting antimitotic effects and for immunosuppression by chemicals. To determine the predictivity of the assay, the lead laboratory first surveyed the literature, generated reference data for immunotoxic profiles of chemicals used in the validation study and the data set. On the other hand, to determine whether chemicals have antimitotic effects or immunosuppression, the lead laboratory made a list of criteria based on the rationale for immunotoxic classification of chemicals proposed by Luster et al. (Luster et al., 1992). Briefly, Luster et al. (1992) presented a panel of immunotoxicity tests and demonstrated that the combination of 3 or more different immunotoxicity tests resulted in 100 % concordance for the judgment of immunotoxicity. The list of criteria we used was based on immunotoxic information and immunotoxicity parameters presented by Luster et al. (1992).

When the performance of the IL-2 Luc LTT was determined with reference to the classification of chemicals with antimitotic effects, it was 76.5 % (13/17) for mean sensitivity, 75.0 % (6/8) for mean specificity, and 76.0 % (19/25) for mean predictivity of the combined data from the Phase I and Phase II trials. When the performance of the IL-2 Luc LTT was determined with reference to the classification of chemicals with immunosuppressive effects, it was 71.4 % (15/21) for mean sensitivity, 100.0 % (4/4) for mean specificity, and 76.0 % (19/25) for mean prediction of the combined data from the Phase I and Phase II trials.

To further characterize the IL-2 Luc LTT assay, the lead laboratory assessed a total of 83 test compounds including 46 pharmaceutical drugs that were composed of 12 anti-cancer drugs, 8 immunosuppressive drugs with antimitotic effects, 8 immunosuppressive drugs without antimitotic effects, 3 JAK inhibitors, and 15 non-immunosuppressive drugs; and 37 non-pharmaceutical chemicals (23 immunosuppressive and 14 non-immunosuppressive). Seven of the 23 immunosuppressive chemicals have antimitotic properties.

When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the IL-2 Luc LTT for antimitotic effects was 63.0 % (17/27) for sensitivity, 90.2 % (46/51) for specificity, and 80.8 % (63/78) for predictivity. When the applicability domain was taken into consideration, the performance of the IL-2 Luc LTT for antimitotic effects was 66.7 % (16/24) for sensitivity, 90.2 % (46/51) for specificity, and 79.5 % (62/78) for predictivity.

We next examined whether the combination of the IL-2 Luc LTT and the IL-2 Luc assay improves the performance of the IL-2 Luc LTT for immunosuppressive effects as part of an integrated approaches to testing and assessment (IATA) approach. When pharmaceutical drugs were examined, the performance of the IL-2 Luc assay alone for immunosuppressive effects was 61.3 % (19/31) for sensitivity, 53.3 % (8/15) for specificity, and 58.7 % (27/46) for predictivity. When non-pharmaceutical chemicals were examined, the performance of the IL-2 Luc assay alone for immunosuppressive effects was 65.2 % (15/23) for sensitivity, 64.3 % (9/14) for specificity, and 64.9 % (24/37) for predictivity. When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the IL-2 Luc assay alone for immunosuppressive effects was 63.0 % (34/54) for sensitivity, 58.6 % (17/29) for specificity, and 61.4 % (51/83) for predictivity.

When pharmaceutical drugs were examined, the performance of the combined IL-2 Luc + IL-2 Luc LTT assay for immunosuppressive effects was 93.3 % (28/30) for sensitivity, 50.0 % (7/14) for specificity, and 79.5 % (35/44) for predictivity. When non-pharmaceutical chemicals were examined, the performance of the combined IL-2 Luc + IL-2 Luc LTT assay for systemic immunosuppressive effects was 82.6 % (19/23) for sensitivity, 58.3 % (7/12) for specificity, and 74.3 % (26/35) for predictivity. When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the combined assay for systemic immunosuppressive effects was 88.7 % (47/53) for sensitivity, 53.8 % (14/26) for specificity, and 77.2 % (61/79) for predictivity. After considering the applicability domain, the performance of the

combined IL-2 Luc + IL-2 Luc LTT assay was 90.0 % (45/50) for sensitivity, 53.8 % (14/26) for specificity, and 77.6% (59/76) for predictivity.

These results suggest that, although the IL-2 Luc LTT alone is not sufficient to detect the immunosuppressive effects of chemicals, the combination with the IL-2 Luc assay can be a promising approach to detect immunosuppression by chemicals. Like the IL-2 Luc assay or the IL-8 Luc assay, chemicals that require metabolic activation or are poorly water-soluble should be outside of the applicability domain. In this study, there were 3 chemicals outside of the applicability domain. Even though these applicability domains are taken into consideration, either the IL-2 Luc assay alone, or in combination with the IL-2 Luc LTT cannot cover all the effects of chemicals on human immune system. Therefore, it is imperative to further 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.

2. Objective and rationale for the study

The objective of the present validation study was to determine the usefulness and limitations of the IL-2 luciferase leukocyte toxicity test (IL-2 Luc LTT) 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 Luc LTT;

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 immunotoxic profiles of the chemicals.

The rationale is multifaceted. The development of this assay is important because:

1) Environmental contaminants, food additives, and drugs can affect the immune system, resulting in immune dysregulation;

2) Immune dysregulation can have serious adverse health consequences, ranging from reduced resistance to infection and neoplasia to allergic and autoimmune conditions;

3) In vitro methods are needed to detect immune effects of chemicals due to the high cost, ethical concerns, and questionable relevance to risk assessment for humans using animal cells and model systems;

4) An IATA approach needs to be developed since immune effects of chemicals are unlikely to be detected by a single in vitro assay;

5) Antimitotic effects of chemicals could not be detected by the IL-2 Luc assay; and

6) The regulatory application is eventual generation of an OECD test method for immunotoxicity.

IL-2 is a reasonable initial target to establish immunotoxicity testing as it is a key cytokine in T cell activation, proliferation, and maintenance of various T cell subsets. As noted, one of the first assays to be validated as part of the larger MITA

was the IL-2 Luc assay, which provided judgement of chemicals to be immunotoxic through alteration (often suppression) of the IL-2 promoter. Surprisingly, many chemicals that are known immunosuppressants (i.e., cyclophosphamide, azathioprine (AZ), mycophenolic acid (MPA), mizoribine (MZR), and methotrexate (MT)) had no effect on IL-2 Luc. It was therefore hypothesized that chemicals that exhibit immunosuppressive effects via antimitotic effects may instead be judged immunosuppressive using the IL-2 Luc LTT assay. It is also proposed that the combination of the IL-2 Luc with IL-2 Luc LTT can be part of an IATA approach to assess the immunotoxic potential of 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. Many chemicals including environmental contaminants, food additives, and pharmaceutical drugs can affect the immune system, resulting in immune dysregulation. Currently, the assessment of chemical immunotoxicity relies mainly on animal models and assays that characterize immunosuppression and sensitization. However, animal trials have many drawbacks, such as high cost, ethical concerns, and questionable relevance and inconsistent translatability to risk assessment for humans (Adler et al.,

2011). In addition, current *in vivo* models do not always provide a mechanistic understanding of the data.

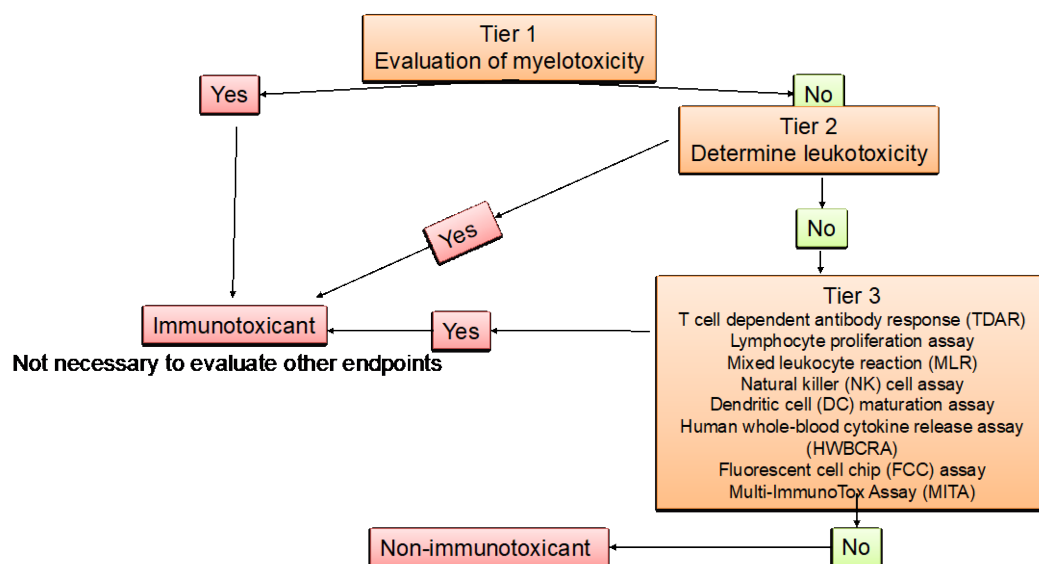
3-2. The current status of *in vitro* immunotoxicity tests

Overcoming some of the drawbacks mentioned above requires the development of *in vitro* methods to detect immunotoxicity. A workshop hosted by the European Centre for the Validation of Alternative Methods in 2003 focused on state-of-the-art *in vitro* systems for evaluating immunotoxicity (Galbiati, Mitjans and Corsini, 2010; Gennari et al., 2005; Lankveld et al., 2010), and among the recommendations, a tiered approach was proposed to assess in vitro immunotoxicity as shown in Fig.1. The proposed tiered approach would begin with pre-screening for direct immunotoxicity by evaluating myelotoxicity (Tier 1). Compounds capable of damaging or destroying bone marrow will most likely have immunotoxic effects. If compounds are not potentially myelotoxic, they are tested for leukotoxicity (Tier 2). Compounds are then tested for immunotoxicity using more mechanistically-specific approaches, such as T cell–dependent antibody response (TDAR), lymphocyte proliferation assay, mixed lymphocyte reaction (MLR), natural killer (NK) cell assay, dendritic cell (DC) maturation assay and human whole-blood cytokine release assay (Tier 3). At present there is no consensus on which assays to use, or how, and there are no OECD test guidelines that describe how to detect chemical immunotoxicity *in vitro*.

Considering the complexity of the immune system and the different underlying mechanisms of immunotoxicity, it is also clear that one assay alone cannot cover all the

potential adverse effects of chemicals on the immune system and a larger set of assays, that will cover the spectrum of immunotoxicity, is needed.

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



This figure is modified from Corsini and Roggen. Overview of in vitro assessment of immunotoxicity DOI: 10.1016/j.cotox.2017.06.016.

3-3. Predictivity of *in vitro* immunotoxicity tests

A crucial step in developing an *in vitro* immunosuppression test involves determining its predictivity. Determining the predictability of tests requires reference chemicals that are positive or negative controls based on results obtained by gold standard analysis or data obtained from the literature. The reference data for the systemic immunosuppression tests vs target specific immunotoxicity tests should be different as they should consider the underlying mechanism of action and immune cells targeted.

In addition, as also applied to non-animal tests for skin sensitization, determining the predictivity of immunosuppression tests cannot be dependent on a single non-animal

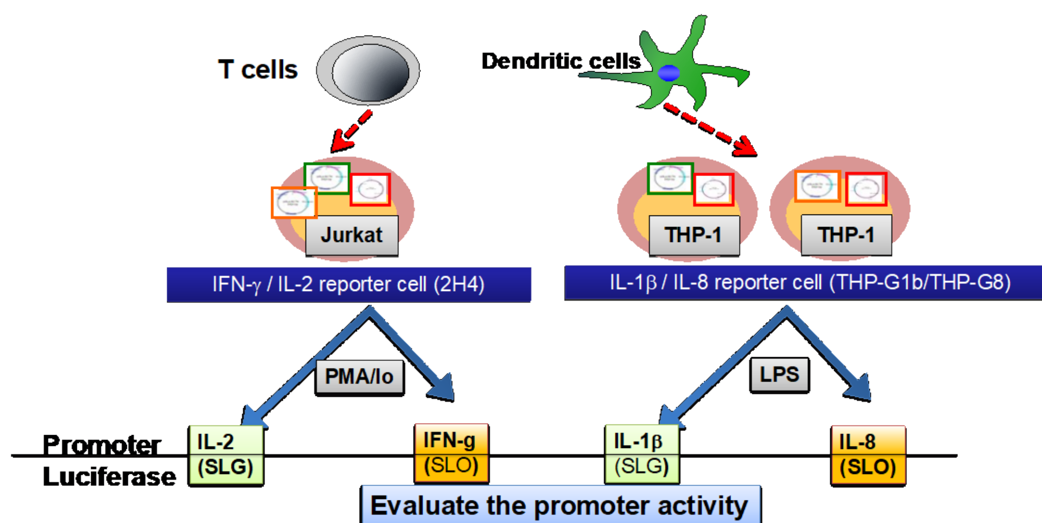
alternative assay; rather, it is necessary to develop an IATA approach. Since systemic immunosuppression likely involves multiple target cells and tissues and/or modes of action, it will need to be addressed by multiple adverse outcome pathways (AOPs). As target-specific effects are more focused, they may be sufficiently explained by a single AOP using combinations of assays representing different key events (KEs) of the AOP. Therefore, it may be difficult to reach the predictability goal for a single *in vitro* immunosuppression test in validation trials because its predictability will depend on the percentage of chemicals affecting the specific readout of the test.

3-4. Multi-ImmunoTox Assay (MITA)

Taking into consideration these limitations, our group developed a high-throughput screening system known as the Multi-ImmunoTox Assay (MITA) to evaluate chemical immunotoxicity as shown in Fig.2. The MITA utilizes three stable reporter cell lines: 1) 2H4 cells derived from Jurkat cells, expressing stable luciferase green (SLG) regulated by the interleukin (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); 2) THP-G8 cells derived from THP-1 cells, expressing SLO regulated by the IL-8 promoter and SLR regulated by the GAPDH promoter (Takahashi et al., 2011); and 3) THP-G1b cells derived from THP-1 cells, expressing SLG regulated by the IL-1β promoter and SLR regulated by the GAPDH promoter (Kimura et al., 2014). These four 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). This assay system can identify the effects of chemicals on IL-2 promotor-driven luciferase activity (IL2LA) and IFN-γ promotor-driven luciferase activity (IFNLA) in 2H4 cells in the presence of the stimulants phorbol 12-myristate 13-acetate (PMA) and ionomycin (Io) and the effects of chemicals on IL-1 promotor-driven luciferase activity (IL1LA) in THP-G1b cells and IL-8 promotor-driven luciferase activity (IL8LA) in THP-G8 cells in the presence of lipopolysaccharide (LPS). We designated the respective components of the MITA as the IL-2 Luc assay for detecting the effects of chemicals on IL2LA; the IFN Luc assay for detecting the effects of chemicals on IFNLA; the IL-1 Luc assay for detecting the effects of chemicals on IL1LA; and the IL-8 Luc assay for detecting the effects of chemicals on IL8LA.

Fig. 2 The Multi-ImmunoTox Assay (MITA)



Takahashi T et al. Toxicol Sci, 2011; Kimura et al. Toxicol in vitro, 2014; Kimura Y et al. Toxicol In Vitro, 2015; Kimura et al. Arch Toxicol. 2018; Kimura et al. J Toxicol Sci, 2018; OECD442E

After establishing the MITA, we first compared the effects of dexamethasone (Dex), cyclosporine (CyA) and tacrolimus (Tac) on the mRNA expression in the three MITA cell lines with the original cell lines, such as Jurkat cells or THP-1 cells, and with human whole-blood cells stimulated with PMA/Io or LPS. The results confirmed that the MITA correctly reflects changes in mRNA expression in the original cell lines and whole-blood cells (Kimura et al., 2014).

3-5. The evaluation of immunotoxicity profiles of well-known immunosuppressive drugs by the MITA

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. Drug classifications are based on a review by Allison (Allison, 2000).

The MITA demonstrated that Dex significantly suppressed IL-2, IL-1β and IL-8 reporter activity, while CyA and Tac suppressed IL-2 and IFN-γ reporter activity but had no effect on IL-1β and IL-8 reporter activity. 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), or 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 Fluorescent Cell Chip (FCC) assay (Wagner et al., 2006). The MITA has the advantage over other assays in that it can discriminate the effects of chemicals on T cells from those on macrophages/dendritic cells.

Table 1. The MITA can detect immunomodulatory 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

S and A indicate drugs that showed statistically significant suppression (S) or augmentation (A) in triplicate experiments for each parameter, while N indicates drugs that did not show significant effects. [1]

3-7. The IL-2 Luc assay plays a principal role in detecting the immunosuppressive effects of chemicals in the MITA

Table 1 indicated that among the assays included in the MITA, the IL-2 Luc assay plays a major role in detecting the immunosuppressive effects of chemicals because the IL-2 Luc assay could detect more numbers of immunosuppressive drugs than any other component of the MITA (Kimura et al., 2014; Kimura et al., 2018; Kimura et al., 2020).

3-8. The IL-2 Luc assay cannot detect immunosuppressive effects of drugs which are dependent on the inhibition of DNA synthesis or anti-proliferative effects

Although the IL-2 Luc assay could detect a significant number of immunosuppressive drugs such as dexamethasone, calcineurin inhibitors, and several off-label immunosuppressive drugs, it could not detect immunosuppressive drugs whose effects are dependent on the inhibition of DNA synthesis or anti-proliferative effects, such as rapamycin (RPM), mizoribine (MZR), cyclophosphamide (CP), methotrexate (MTX), and mycophonolic acid (MPA) (Kimura et al., 2014; Kimura et al., 2018; Kimura et al., 2020).

[1] Table 1 was adapted from Kimura et al., 2014

3-9. The IL-2 Luc LTT can detect antimitotic effects of chemicals.

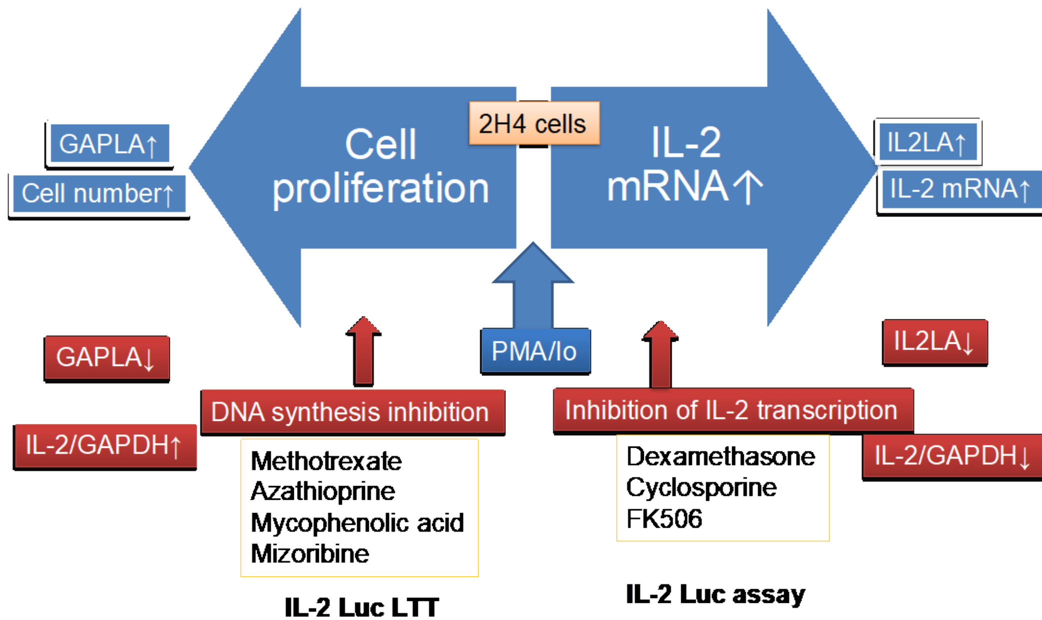
Recently, Shao et al. demonstrated that a toxicogenomics approach using Jurkat cell lines can identify mechanisms underlying the direct immunotoxicity of chemicals, such as endoplasmic reticulum stress, response to oxidative stress, regulation of cell cycle, and anti-apoptosis (Shao et al., 2014). We therefore speculated whether an assay could be developed to detect the antimitotc effects of chemicals using 2H4 cell lines.

The 2H4 cell lines simultaneously measure GAPDH promoter-driven luciferase activity (GAPLA), IL-2LA, and IFNLA (Saito et al., 2011). GAPDH mRNA is ubiquitously expressed at moderately abundant levels. It is frequently used as an endogenous control for quantitative real time polymerase chain reaction because its expression is constant at different times and after experimental manipulation in some experimental systems (Edwards and Denhardt, 1985; Mori et al., 2008; Winer et al., 1999). Although there are several reports suggesting that its use as an internal standard is inappropriate in some cases (Oliveira et al., 1999; Thellin et al., 1999), in general, there is little within-tissue variation of GAPDH mRNA expression levels (Barber et al., 2005). In addition, in THP-G8 cells that contain the same plasmid as 2H4 (which contains the SLR gene driven by the GAPDH promoter), we previously reported a significant correlation between Inh-GAPLA that is defined as GAPLA of reporter cells treated with chemicals/GAPLA of untreated cells and the reduction of propidium iodide (PI)-excluding cells, with strengthened correlation with cytotoxicity depending on the culture conditions. In other words, GAPLA shows a good correlation with viable cell number and can be used as a marker of cell viability, and in addition, a decrease in

GAPLA precedes a decrease in PI-excluding cells (Kimura et al., 2015; Takahashi et al., 2011).

We therefore hypothesized that if 2H4 cells are treated with antimitotic agents for 24 h and then stimulated with PMA/Io, they may reduce GAPLA without suppressing either IL2LA or IFNLA (Fig. 3). Therefore, we modified the protocol of the IL-2 Luc assay by changing the incubation period with chemicals from 1 h to 24 h and seeding cell density. Then, we evaluated the performance of the assay using pharmaceutical drugs with well-known effects on the immune system. We first examined CyA, AZ, MPA, MZ, bleomycin, and the control dimethyl sulfoxide (DMSO) by the IL-2 Luc assay with the modified protocol. As expected, CyA did not affect GAPLA and suppressed the ratio of IL2LA/GAPLA, while AZ, MPA, MZ, and bleomycin significantly suppressed GAPLA and increased the ratio of IL2LA/GAPLA. After we demonstrated that the results of the luciferase assay corresponded well with the actual change in cell number and IL-2 production, we constructed an assay to detect the antimitotic effects of chemicals using 2H4 cells by more precisely defining the protocol and criteria to judge positive or negative. The IL-2 luciferase assay with this modified protocol is designated as the IL-2 luciferase leukocyte toxicity test (IL-2 Luc LTT).

Fig. 3. The different aspects of immunosuppression detected by the IL-2 Luc assay and the IL-2 Luc LTT



3-10. The process of validation of the combined assays of the IL-2 Luc assay and the IL-2 Luc LTT

We have already conducted the validation study of the IL-2 Luc assay (Kimura et al., 2020). In this study, we conducted the validation study for the IL-2 Luc LTT. First, we demonstrated the within- and between-laboratory reproducibility and then calculated its performance for detecting anti-mitotic effects and immunosuppressive effects. In addition, we also evaluated the performance of the combination of the IL-2 Luc assay and the IL-2 Luc LTT as part of an IATA approach.

4. Test method and modification

4-1. IL-2 reporter cell line 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™ 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 bioluminescent substrate. Namely, 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 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

The 2H4 cells (1×10^4 cells/50 μL/well) in a 96-well black plate were pretreated with different concentrations of individual chemicals for 24 hours. Next, 2H4 cells were

stimulated with 25 nM of Phorbol 12-myristate 13-acetate (PMA) and 1 mM of Ionomycin for 6 hours. Two luciferase activities (SLG luciferase activity (SLG-LA) and SLR luciferase activity (SLR-LA)) were simultaneously determined using a microplate-type luminometer with a multi-color detection system (e.g., Phelios AB-2350 (ATTO), ARVO (PerkinElmer), Tristar LB941 (Berthold)) and Tripluc Luciferase Assay Reagent (TOYOBO Co., Ltd., Osaka, Japan) according to the manufacturers' instructions. We obtained SLG-LA driven by the IL-2 promoter (IL2LA) and SLR-LA driven by GAPDH promoter (GAPLA) in 2H4 cells. 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 2H4 cells. In addition, we calculated % suppression and Inh-GAPLA as follows:

% suppression= (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 LTT

Abbreviations	Definition
GAPLA	SLR luciferase activity reflecting GAPDH promoter activity
IL2LA	SLG luciferase activity reflecting IL-2 promoter activity of 2H4 cells
IFNLA	SLO luciferase activity reflecting IFN-g promoter activity of 2H4 cells
nIL2LA	IL2LA/GAPLA of 2H4 cells
nIFNLA	IFNLA/GAPLA of 2H4 cells
% suppression	(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
CV05	The lowest concentration of the chemical at which Inh-GAPLA becomes < 0.05
Min Inh-GAPLA	The minimum value of Inh-GAPLA of each experiment

4-3. Criteria to determine antimitotic effects of chemicals

We set the acceptance criteria and criteria for the Phase I trial of the IL-2 Luc LTT as follows (IL-2 Luc LTT protocol ver. 1.3):

Acceptance criteria

At the time of each experiment, a control experiment examining nIFNLA of 2H4 cells treated with PMA/Io and nIFNLA of non-treated 2H4 cells must be conducted. Then, the fold induction of nIFNLA of 2H4 cells treated with PMA/Ionomycin to nIFNLA 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.

Criteria

The experiments are repeated until 2 consistent antimitotic results, indeterminate results, or non-antimitotic results are obtained. When 2 consistent results are obtained, the chemicals are judged as indicated by the obtained consistent results.

In each experiment, if chemicals meet the following criteria described below and give Min Inh-GAPLA < 0.7, they are judged as antimitotic. Otherwise, they are judged as provisional non-antimitotic.

The criteria for stimulatory:

1. The mean of % suppression is ≤-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 stimulatory data points or 1 statistically significant stimulatory data point with a trend in which at least 3 consecutive data points decrease in a dose-dependent manner. In the latter case, the trend can cross 0, if 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.

Of chemicals that are not judged as provisional non-antimitotic, if chemicals do not demonstrate statistically significant suppressive or stimulatory data points, show Min Inh-GAPLA ≥ 0.7, and they are insoluble at 10 mg/mL in distilled water, they are judged as indeterminate because they may be not dissolved in the vehicle at the concentration sufficient to show the effects in the culture medium. Otherwise, they are judged as non-antimitotic.

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 a virus promoter (e.g., the herpes simplex virus thymidine kinase promoter, simian virus 40 promoter) or a housekeeping gene promoter (e.g., GAPDH, β-actin). This assay system has been commercialized as the Dual-Luciferase Reporter Assay System by Promega Corporation. In this system, both luciferase activities are measured sequentially from single extracts based on their bioluminescent substrate specificity. Firefly luciferase activity is measured first by adding firefly D-luciferin, 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 preferable to use an improved assay system whereby gene expression can be monitored simultaneously. In the MITA, therefore, three kinds of beetle luciferases that emit either green, orange, or red light with a single bioluminescent substrate, D-luciferin, are used. Multiple promoter activities are conventionally evaluated in a one-step reaction by combined use of a commercially available bioluminescent reagent (Tripluc Luciferase Assay Reagent, TOYOBO) and a microplate luminometer equipped with optical filters (Nakajima et al. 2005, 2010).

In the IL-2 Luc LTT, the triple-color assay system consists of a green-emitting luciferase (SLG; λmax = 550 nm) (Ohmiya et al. 2000; Nakajima et al. 2005) for monitoring IL-2 promoter activity, an orange-emitting luciferase (SLO; λmax = 580 nm) (Viviani et al. 2001; Nakajima et al. 2005) for monitoring IFN-γ promoter activity, and a red-emitting luciferase (SLR; λmax = 630 nm) (Viviani et al. 1999; Nakajima et al. 2005) for monitoring internal control promoter (GAPDH) activity. 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, measured using optical filters (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 of light that passed through the O56 filter (>560-nm long-pass filter) or the R60 filter (>600-nm long-pass filter), respectively, are 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_{O56}, κO_{O56} and κR_{O56} are the transmission coefficients for the green-, orange- and red-emitting luciferases of the O56 filter, respectively, and κG_{R60}, κO_{R60} and κR_{R60} are the transmission coefficients for the green-, orange- and red-emitting luciferases of the R60 filter, respectively. The transmission coefficients are simply estimated using purified recombinant luciferase enzymes (Niwa et al. 2010).

$$\begin{pmatrix} F0 \\ F1 \\ F2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ \kappa G_{O56} & \kappa O_{O56} & \kappa R_{O56} \\ \kappa G_{R60} & \kappa O_{R60} & \kappa R_{R60} \end{pmatrix} \begin{pmatrix} G \\ O \\ R \end{pmatrix}$$

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).

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 (Retired, Tohoku University, Miyagi, Japan), Developer of this assay, Test method, expertise underlying science

Yutaka Kimura (Retired, Tohoku University, Miyagi, Japan)

International expert members

ECVAM 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 (Division of Translational Toxicology (DTT)/NIEHS, USA), Immunotoxicity expertise

JSIT liaison: Tomoaki Inoue (Retired, Chugai Pharmaceutical Co., Ltd.), Immunotoxicity expertise

Data management team: Takashi Omori (Department of Clinical Biostatistics Graduate School of Medicine, Kyoto University Japan), Data analysis, biostatistics dossier

Chemical Selection Committee（CSC）:

Setsuya Aiba (Retired, Tohoku University)
Yutaka Kimura (Retired, Tohoku University)
Hajime Kojima (JaCVAM)
Emanuela Corsini (Milan Univ)

Erwin L. Roggen (3Rs Management and Consulting ApS)
Dori Germolec DTT/NIEHS)
Tomoaki Inoue (Retired, Chugai Pharmaceutical Co., Ltd.)

Participating Test Facilities

Test Facility 1: Tohoku University, SD: Retired, Chizu Fujimura
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 Tonomachi Kawasaki, Kawasaki, 210-9501
TEL: +81-44-270-6600
h-kojima@nihs.go.jp

5-3. Meetings held

June 12th, 2020 (Web meeting)
The teleconference for the MITA Validation study
Subjects: Proposal of new protocol and future plan for IL-2 Luc assay LTT
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T., Aiba, S., Kimura, Y., Omori, T., Nakajima, Y. Yasuno, R., Kojima, H.

July 21st, 2020 (Web meeting)
The teleconference for the MITA Validation study
Subjects: Proposal of new protocol and study plan for IL-2 Luc assay LTT
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T., Aiba, S., Kimura, Y., Omori, T., Nakajima, Y., Yasuno, R., Kojima, H.

September 8th, 2020 (Web meeting)
The teleconference for the MITA Validation study

Subjects: Results of phase 0 in IL-2 Luc LTT and discussion for protocol.
In house data for IL-2 Luc LTT
Chemical selection for phase I (VMT only)
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T., Aiba, S., Kimura, Y., Nakajima, Y., Yasuno, R., Kojima, H.

January 7th, 2021 (Web meeting)
The teleconference for the MITA Validation study
Subjects: Results of phase 1 in IL-2 Luc LTT
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T., Aiba, S., Kimura, Y., Omori, T., Nakajima, Y., Yasuno, R., Kojima, H.

July 8th, 2021 (Web meeting)
The teleconference for the MITA Validation study
Subjects: Results of phase 2 in IL-2 Luc LTT
VMT members: Corsini, E., Roggen, E., Germolec, D., Inoue, T., Aiba, S., Kimura, Y., Omori, T., Nakajima, Y., Yasuno, R., Kojima, H.

6. Study Design

The validation trial assessed the reliability (reproducibility within and between laboratories) and relevance (predictive capacity) of the IL-2 Luc LTT with a challenging set of test substances (test items) for which high quality *in vitro and in vivo* data are available. The validation study (Phase I and Phase II trials) was conducted by three laboratories, based on the study design in accordance with study plan (Appendix 1) and using the test chemicals shown in Table 3. The methods were described above in Section 4 and the precise protocol is described later in Section 8.

7. Test Chemicals

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

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

Table 3. Breakdown of the IL-2 Luc LTT validation study

Trial	Number of test compounds	Number of repetitions	Information obtained	Experiment date
Phase 0	3 non-coded	1	Between-lab transferability	August, 2020
Phase I	5 coded	3	Within & between-lab reproducibility	September to December, 2020
Phase II	20 coded	1	Between-lab reproducibility & predictability	May to August, 2021

7-1. Basic rules for chemical selection

The selection of test chemicals by the CSC in the VMT was based on published papers on *in vivo* immunotoxicity tests and validation trials for *in vitro* alternative assays on immunotoxicity test methods.

7-1-1. Applied selection criteria

- information on mode/site of action such as inhibition of cytokine production and antimitotic effects
- 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 trials
- 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 implied 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, and an EPA candidate list. An extensive literature search was performed by the CSC 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) 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 (Appendixes 2-1 and 2-2), in collaboration with the CSC. The 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 under conditions in accordance with the storage instructions and received sealed safety information such as the Material Safety Data Sheet (MSDS) describing hazard identification and exposure control/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 validation study, and all test facilities returned the MSDSs for the test chemicals to JaCVAM in a 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 - Phase 0 trial

Transferability of this assay was checked using three non-coded chemicals: bleomycin sulfate, dexamethasone and 6-thioguanine (Appendix 3) in three 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 five coded chemicals in three test facilities, as shown in Table 4 and Appendix 4. These chemicals were selected by the CSC based on the in-house dataset of the lead laboratory. The chemicals were coded by JaCVAM as shown in Table 4 and distributed to the test facilities. Bleomycin sulfate and cyclosporin were used as positive and negative controls, respectively.

Table 4. Chemical code list on the phase I trial for IL-2 Luc LTT

No.	Chemical name	CAS No.	Storage	Physicality	Supplier	Lot
1	Mycophenolic acid	24280-93-1	0～10℃	Solid	TCI	CUHFB-QK
2	Indomethacin	53-86-1	Room T.	Solid	SIGMA	059M4169V
3	Cyclosporin A	59865-13-3	2～10℃ (No Freezing)	Solid	WAKO	15-XJZ-47-1

4	5-FU	51-21-8	Room T.	Solid	WAKO	CAH4560
5	Mannitol	69-65-8	Room T.	Solid	SIGMA	SLCD7105

7-4. Validation study - Phase II trial

Twenty test chemicals were selected by CSC for between-laboratory reproducibility as shown in Table 5 and Appendix 5. The chemicals were coded by JaCVAM as shown in Table 5 and distributed to the test facilities.

Table 5. Chemical code list on the phase II trial for IL-2 Luc LTT

No.	Chemical name	CAS No.	Storage	Physicality	Supplier	Lot	Remark
1	Colchicine	64-86-8	Room T.	Solid	SIGMA	SLCB8521	
2	Tetrabromobisphenol A EP	79-94-7	Room T.	Solid	TCI	MHI2H-HT	
3	Fludarabine	21679-14-1	Room T.	Solid	TCI	QP38F-ER	
4	Nicotinamide (NA)	98-92-0	Room T.	Solid	SIGMA	BCCC5163	
5	Glycidol	556-52-5	2～8℃	Liquid	SIGMA	BCCB9452	Dangerous Substance
6	Perfluorooctanoic Acid	335-67-1	Room T.	Solid	TCI	HDBRI-FJ	
7	Citral	5392-40-5	Room T.	Liquid	SIGMA	STBJ0146	
8	Cytarabine EP	147-94-4	Room T.	Solid	TCI	O6YGN-AF	
9	Etoposide	33419-42-0	Room T.	Solid	TCI	AADNB-PF	
10	Gemcitabine hydrochloride	122111-03-9	Room T.	Solid	TCI	P4UDM-DD	
11	6-Thioguanine	154-42-7	Room T.	Solid	TCI	2H2WO-QG	
12	Cadmium chloride	10108-64-2	Room T.	Solid	Wako	PEE3332	Deleterious Substance
13	Prednisolone	50-24-8	Room T.	Solid	SIGMA	SLCB2470	
14	Azathioprine (AZ)	446-86-6	-20℃	Solid	SIGMA	MKCM1683	
15	2-Methyoxyacetic Acid	625-45-6	Room T.	Liquid	SIGMA	STBD2084V	Dangerous Substance
16	4-Chloro-o-phenylenediamine	95-83-0	Room T.	Solid	SIGMA	14606EDV	
17	Dichloroacetic acid	79-43-6	2～8℃	Liquid	SIGMA	SHBK0353	Deleterious Substance
18	n-Nitrosodimethylamine	62-75-9	Room T.	Liquid	TCI	VVOVL-BS	Dangerous Substance
19	Methyl carbamate	598-55-0	Room T.	Solid	TCI	4P6SO-NB	
20	Cyclophosphamide monohydrate	6055-19-2	2～8℃	Solid	SIGMA	MKCL-2547	

7-5. Acceptance criteria

The within-laboratory reproducibility for all the test facilities was determined by independent biostatistical analysis using five coded chemicals, under supervision by the VMT in the Phase I trial. The proportion of concordance for within-laboratory reproducibility should be greater than or equal to 80% to be accepted as tentative acceptance criteria.

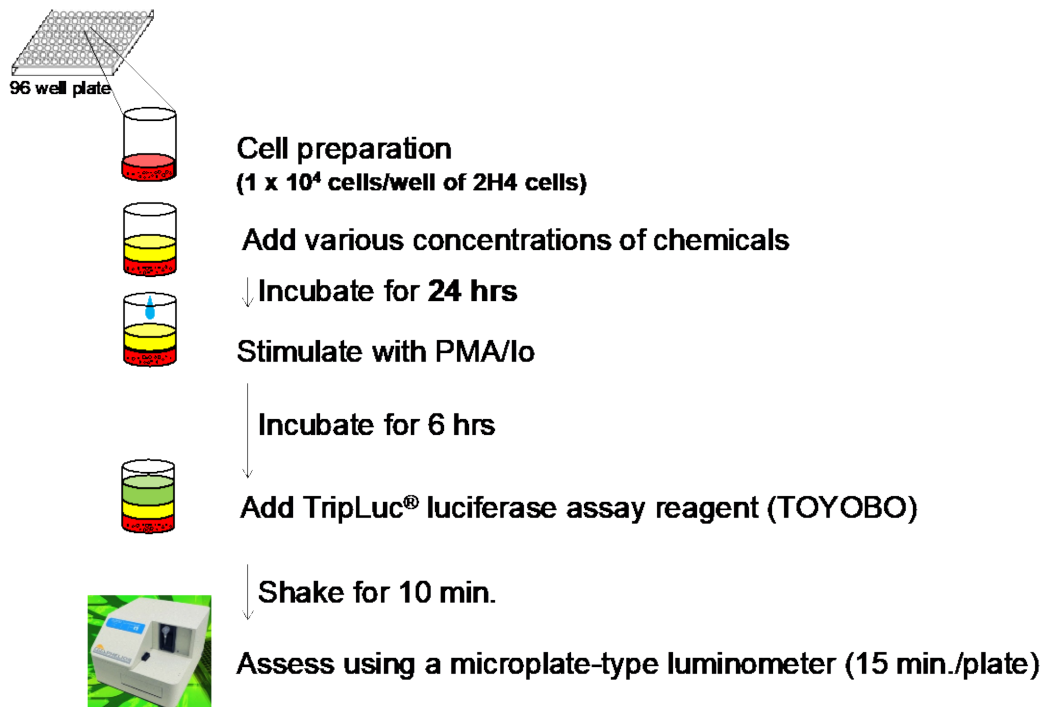
Twenty-five coded test items were selected to confirm the between-laboratory reproducibility in the Phase I and II trials. At the end of the testing, the test facilities submitted a QC certified copy of the whole study dossier to the trial coordinator (study plan, raw data, records and data analysis, study report in accordance with the principles of GLP [OECD, 2018]). The proportion of concordance for between-laboratory reproducibility should be greater than or equal to 80% to be accepted as acceptance criteria.

8. Protocols

8-1. Overview of IL-2 Luc LTT

An overview of the IL-2 Luc LTT is shown in Fig. 4. In addition, the final protocol for the present test (version 001.8) is provided as Appendix 6, 7, and 8 and the procedures are described in detail below.

Fig. 4. Overview of the IL-2 Luc LTT



8-2. Protocol for IL-2 Luc LTT

8-2-1. Reagents and equipment

The following reagents and equipment were used.

For maintenance of the 2H4 cells

- RPMI-1640 (GIBCO Cat#11875-093, 500 mL)
- FBS (Biological Industries Cat#04-001-1E Lot: 1524129)
- 100× concentrated antibiotic and antimycotic (10,000 U/mL of penicillin G, 10,000 mg/mL of streptomycin and 25 mg/mL of amphotericin B in 0.85% saline) (e.g., GIBCO Cat#15240-062)
- Hygromycin B (CAS:31282-04-9, Invitrogen Cat#10687-010)
- G418 (CAS:108321-42-2, WAKO Cat#074-06801)
- 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

- Tripluc® Luciferase Assay Reagent (TOYOBO Cat#MRA-301)

Expendable supplies

- T-75 flask tissue culture treated (e.g., Corning Cat#353136)
- 96 well clear black plate (flat-bottom, for measurement of the luciferase activity, e.g., Greiner Bio-one Cat#655090)
- 96 well clear plate (round-bottom, for preparation of chemicals and stimulants)
- 96 well assay block, 2 mL (e.g., Costar Cat#3960)
- Seal for 96 well plate (e.g., Perkin Elmer TopSeal-A PLUS Cat#6050185, EXCEL Scientific SealMate Cat#SM-KIT-SP)
- Reservoir
- Pipette

Equipment for measurement of luciferase activity

- Measuring device: a microplate-type luminometer with a multi-color detection system that can accept two optical filters
  e.g., Phelios AB-2350 (ATTO), ARVO (PerkinElmer), Tristar LB941 (Berthold)
- Optical filter: 560 nm long-pass filter and 600 nm long-pass filter

Others

- Pipetman
- 8 channel or 12 channel pipetman (optimized for 10-100 μL)
- Plate shaker (for 96-well plate)

- CO_2 incubator (37°C, 5% CO_2)
- Water bath
- Cell counter: hemocytometer, trypan blue

8-2-2. Culture media

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

Table 6. 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 Cat #11875-093	-	-	440 mL
FBS	Biological Industries Cat#04-001-1E Lot: 1524129	-	10 %	50 mL
Antibiotic-Antimycotic	e.g., GIBCO Cat #15240-062	100×	1×	5 mL
Puromycin	InvivoGen # ant-pr-1	10 mg/mL	0.15 μg/mL	7.5 μL
G418	WAKO Cat#074-06801	50 mg/mL	300 μg/mL	3 mL
HygromycinB	Invitrogen #10687-010	50 mg/mL	200 μg/mL	2 mL

Table 7. B medium: for luciferase assay (30 mL, stored at 2-8°C)

Reagent	Company	Concentration	Final concentration in medium	Required amount
RPMI-1640	GIBCO Cat #11875-093	-	-	27 mL

FBS	Biological Industries Cat#04-001-1E Lot: 715004	-	10 %	3 mL

Table 8. 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 Cat #11875-093	-	-	26.7 mL
FBS	Biological Industries Cat#04-001-1E Lot: 1524129	-	10 %	3 mL
Antibiotic-Antimycotic	e.g., GIBCO Cat #15240-062	100×	1×	0.3 mL

8-2-3. Preparation of stimulant for 2H4 cells

Table 9-1. Phorbol 12-myristate 13-acetate (PMA)

Reagent	Company	Concentration of stock solution	Final concentration
Phorbol 12-myristate 13-acetate (PMA)	Sigma #P8139	2 mM	25 nM
DMSO	Sigma #D5789		

Dissolve 1 mg PMA using DMSO 811 μL, dispend at 5 μL/tube and store at freezer at -30°C. Use these stocks within 6 months after dissolution.

Table 9-2. Ionomycin

Reagent	Company	Concentration of the stock solution	Final concentration
Ionomycin	Sigma # I0634	2 mM	1 µM
Ethanol	Wako #057-00456		

Dissolve 1mg Ionomycin using ethanol 669.3 µL, dispend at 15 µL/tube and store at freezer at -30°C. Use these stocks within 6 months after dissolution.

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 A medium in a T-75 flask at 37°C in a 5% CO_2 incubator (for culture).

Thaw frozen cells (5×10^6 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 × *g* at room temperature for 5 min, discard the supernatant, and resuspend the cells 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 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 $1\sim3 \times 10^5$/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

Cells should be passaged 3-4 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 (1.0×10^6 cells for two chemicals are required, but to have some leeway, 1.5×10^6 cells for two chemicals should be prepared), centrifuge the tube at 120-350 x *g*, 5 min. Resuspend in the pre-warmed B medium at a cell density of 2×10^5/mL. Transfer the cell suspension to a reservoir (Thermo Scientific) and add 50 µL of cell suspension to each well of a 96-well clear black plate (flat bottom) using an 8 channel or 12 channel Pipetman™ (Gison, Inc., Middleton, WI, USA). Each well of the 96-well plate should contain 1×10^4 2H4 cells + 50 µL of B medium.

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

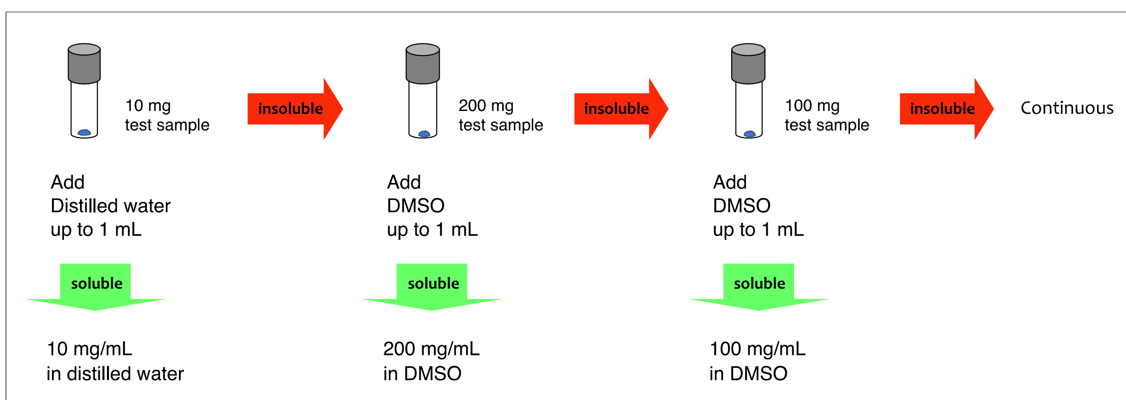
Dissolve the chemical first in distilled water. Weigh 10 mg of the test chemical in a volumetric flask and add distilled water up to 1 mL. If the chemical is soluble at 10 mg/mL, use 10 mg/mL solution for the stock solution.

If the chemical is not soluble at 10 mg/mL in water, the chemical should be dissolved in DMSO at 200 mg/mL. For example, weigh 200 mg of the test chemical in volumetric flask and add DMSO up to 1 mL. If the chemical does not dissolve in DMSO at 200 mg/mL, use the highest concentration possible after diluting with DMSO at a dilution factor of 2.

For chemicals that are difficult to obtain in sufficient quantities, prepare the highest concentration possible instead of 10 mg/mL distilled water. If the chemical is not soluble at 10 mg/mL, prepare the highest concentration possible in DMSO.

Sonication and vortexing may be used if needed, with an attempt to dissolve the chemical for at least 5 minutes. Solubility should be confirmed by the absence of precipitation following centrifugation at 15,000 rpm (≈20,000 x g) for 5 min. The chemical should be used within 4 hours after being dissolved in distilled water or DMSO. (Fig. 5).

Fig. 5. Dissolution by solvent



8-2-8. Dilution of chemicals

For water soluble chemicals, 10 serial dilutions were conducted using distilled water, diluting by a factor of 2, in the 1st, 2nd, 3rd, and 4th experiment. For water insoluble chemicals, 10 serial dilutions were conducted using DMSO as the solvent, diluting by a factor of 2 in the 1st, 2nd, 3rd, and 4th experiments. The solutions are further diluted and added to 2H4 cells in a 96 well plate. After 24-hour incubation at 37°C in a 5% CO_2 incubator, 2H4 cells were added to 10 μL of PMA/ionomycin 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/ionomycin for 6 hours 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 pipette and the plate is shaken for 10 minutes at room temperature 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 seconds 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 9). IL2LA, IFNLA and GAPLA are calculated for each well based on the algorithm to calculate IL2LA, IFNLA and GAPLA from the raw luminescence data reported previously (Nakajima et al., 2005; Noguchi et al., 2008). In addition to being used to calculate IL2LA, IFNLA and GAPLA, this data sheet can automatically generate final

graphs showing the correlation between % suppression and the concentration of chemicals, and between Inh-GAPLA and the concentration of the chemical.

8-2-10. Luminometer apparatus

Multi-color detection systems such as microplate luminometers are available (e.g., Phelios (ATTO, Tokyo, Japan), Tristan 941 (Berthold, Bad Wildbad, Germany), and the ARVO series (PerkinElmer, Waltham, MA)). The luminometer detectors must have high sensitivity (especially for the red region) and low background noise, and are usually equipped with optical filters such as sharp-cut (long-pass) filters or band-pass filters. The transmission coefficients for these filters against each luciferase must be estimated prior to initiating the experiments because the coefficients are dependent on the luminometer due to lot-to-lot variations in detectors.

8-2-11. Positive and negative controls

In each experimental set, bleomycin sulfate and dexamethasone are used as positive and negative controls, respectively.

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

In the IL-2 Luc LTT, nIL2LA was defined to represent IL-2 promoter-driven SLG luciferase activity (IL2LA) normalized by the SLR luciferase activity (GAPLA). The inhibition of GAPLA (Inh-GAPLA) was determined by dividing the GAPLA for 2H4

treated with chemicals with the GAPLA for non-treated 2H4. Percent suppression reflects the effect of chemicals on the IL-2 promoter (Table 10).

Table 10. Abbreviations used in the IL-2 Luc LTT protocol

Abbreviations	Definition
GAPLA	SLR luciferase activity reflecting GAPDH promoter activity
IL2LA	SLG luciferase activity reflecting IL-2 promoter activity of 2H4 cells
IFNLA	SLO luciferase activity reflecting IFN-γ promoter activity of 2H4 cells
nIL2LA	IL2LA/GAPLA of 2H4 cells
nIFNLA	IFNLA/GAPLA of 2H4 cells
% suppression	(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
CV05	The lowest concentration of the chemical at which Inh-GAPLA becomes < 0.05
Min Inh-GAPLA	The minimum value of Inh-GAPLA of each experiment

8-2-13. Acceptance criteria

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

- For each set of experiments, a control experiment examining nIFNLA of 2H4 cells treated with PMA/Io and nIFNLA of non-treated 2H4 cells must be conducted. Then, the fold induction of nIFNLA of 2H4 cells treated with PMA/Ionomycin to nIFNLA

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.

8-2-14. Criteria

The experiments are repeated until 2 consistent antimitotic results, indeterminate results, or non-antimitotic results are obtained. When 2 consistent results are obtained, the chemicals are judged as indicated by the obtained consistent results.

In each experiment, if chemicals meet the following criteria described below and give Min Inh-GAPLA < 0.7, they are judged as antimitotic. Otherwise, they are judged as provisional non-antimitotic (Fig. 6).

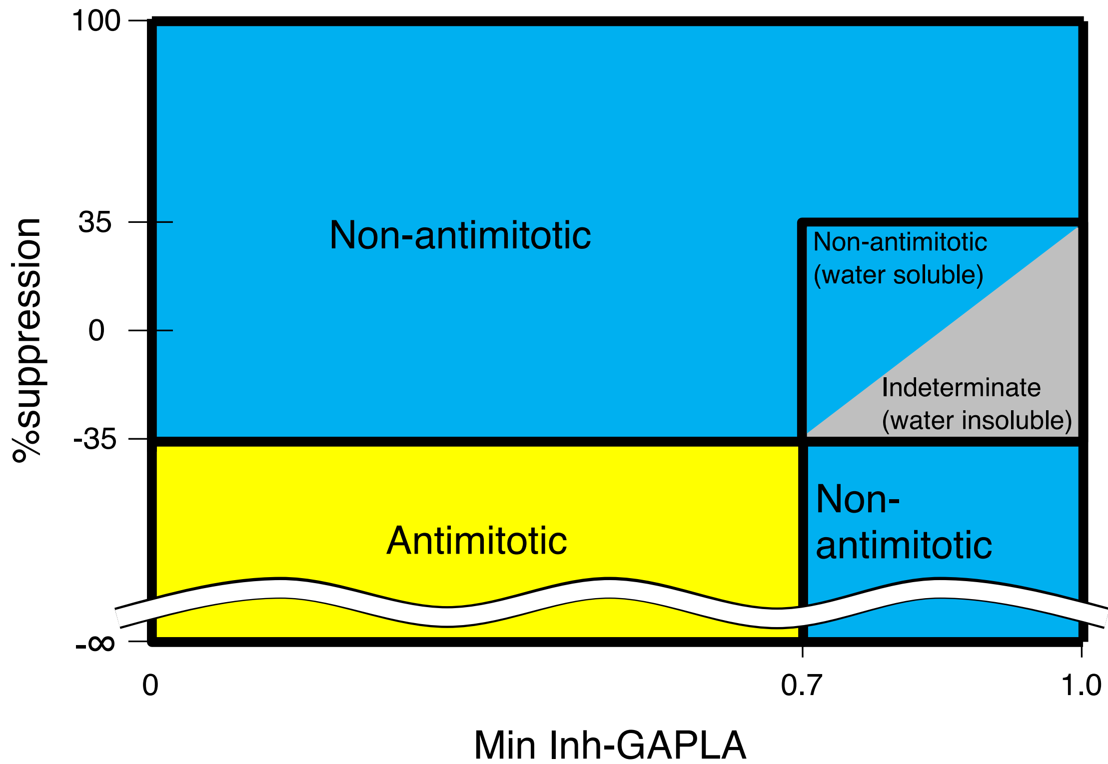
The criteria for antimitotic:

1. The mean of % suppression is ≤-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 stimulatory data points or 1 statistically significant stimulatory data point with a trend in which at least 3 consecutive data points decrease in a concentration-dependent manner. In the latter case, the trend can cross 0, if 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.

Of chemicals that are judged as provisional non-antimitotic, if chemicals do not give statistically significant suppressive or stimulatory data points, give Min Inh-GAPLA ≥ 0.7, and they are insoluble at 10 mg/mL in distilled water, they are judged as indeterminate because they may be not dissolved in the vehicle at the concentration

sufficient to show the effects in the culture medium. Otherwise, they are judged as non-antimitotic.

Fig. 6 Criteria of IL-2 Luc LTT



8-3. Data collection

8-3-1. Operating procedure

Details of the operating procedure for this assay are described in protocol version 001.4 (Appendix 6). Versions of the protocols were updated during the validation trials, but the descriptions of the operating procedures described in these protocols were the same for the two validation trials.

8-3-2. Chemicals

For the Phase I trial, in which the main aim was to evaluate intra- and inter-laboratory reliability, a total of 15 coded chemicals were tested in three rounds of five chemicals each distributed to three laboratories. Different codes were used in the rounds and thus the technician in each laboratory could not identify the chemicals. For the Phase II trial, in which the main aim was to evaluate inter-laboratory reliability, 20 coded chemicals were distributed.

In this document, the chemicals were re-coded. The round is indicated by a suffix such as P101_R1 for the first chemical of the first round of the Phase I trial: P1 means Phase I; 01 means the first chemical; _R1 means first round.

8-3-3. Data handling

The Excel data sheet developed for this study was distributed to the laboratories. Data files were received from each of the three laboratories.

JaCVAM, provided the files listing the chemical codes for the five distributed chemicals in the Phase I trial, and 20 chemicals in the Phase II trial.

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

Since Excel data sheets can display a concentration-response plot for % suppression with its 95% confidence interval, we could judge “Stimulatory” or “Negative” for each experiment from the plot (Appendix 9).

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 (i = 0 to 11) was measured for IL2LA and GAPLA. 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

% suppression is an index for the averaged nIL2LA for the repetition using the i-th concentration compared with at 0 concentration and is the primary measure in this assay. Suppression (%) is described by the following formula:

$$\% \text{ Suppression}_{\mathrm{i}} = \left\{ 1 - \frac{\left(\frac{1}{4}\right) \sum_{\mathrm{i}} \mathrm{nIL2LA_{ij}}}{\left(\frac{1}{4}\right) nIL2\mathrm{LA_{0j}}} \right\} \times 100 \qquad (1)$$

The lead laboratory has proposed that -35 of the value suggests stimulatory for a tested chemical. This value is based on the investigation of the historical data of the lead laboratory. Data management team used this value through all the phases of the present validation study.

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

The lower limit of the 95% CI above 0 is interpreted as the nIL2LA of the i-th concentration being statistically significantly greater than at 0 concentration.

There are several ways to construct the 95% CI. We used the method known as the Delta method. This 95% confidence interval is obtained from the following formula:

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

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 the 97.5 percentile of the standard normal distribution.

8-3-4-2. Inh-GAPLA

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

$$\mathrm{Inh} - \mathrm{GAPLA}_\mathrm{i} = \left(\frac{1}{4}\right) \sum_\mathrm{i} \mathrm{GAPLA}_\mathrm{ij} / \left(\frac{1}{4}\right) \sum_\mathrm{i} \mathrm{GAPLA}_\mathrm{0j}.$$

Since GAPLA is the denominator of the nIL2LA, an extremely small GAPLA value causes a large variation in nIL2LA. Therefore, the i-th % suppression value with an extremely small value of Inh-GAPLA might result in poor precision.

8-3-4-3. Judgment for “antimitotic” or “non-antimitotic” in each experiment

In each experiment, if chemicals meet the criteria described below and give Min Inh-GAPLA < 0.7, they are judged as antimitotic. Otherwise, they are judged as provisional non-antimitotic (Fig. 6).

The criteria for stimulatory:

1. The mean of % suppression is ≤-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 stimulatory data points or 1 statistically significant stimulatory data point with a trend in which at least 3 consecutive data points decrease in a concentration-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.

   Of chemicals that are judged as provisional non-antimitotic, if chemicals do not demonstrate statistically significant suppressive or stimulatory data points, show Min Inh-GAPLA ≥ 0.7, and they are insoluble at 10 mg/mL in distilled water, they are judged as indeterminate because they may be not dissolved in the vehicle at the concentration sufficient to show the effects in the culture medium. Otherwise, they are judged as non-antimitotic.

8-3-4-4. Final judgment for “antimitotic” or “antimitotic” using this assay

In this assay, “antimitotic” or “non-antimitotic” is defined as the case in which there are two identical judgments in a set of experiments.

8-3-5. Reliability

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

Within-laboratory reproducibility was determined by whether or not three experimental sets for each chemical by each laboratory were concordant. The concordance rate was then calculated as the proportion of concordance for each laboratory.

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

To summarize, the concordance rate for within-laboratory reproducibility from the three laboratories was 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 three laboratories for 25 chemicals, this is, five chemicals in the Phase I trial and 20 chemicals in the Phase II trial. These judgements were tabulated, and 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 three laboratories was used to calculate the averaged concordance rate.

8-3-6. Predictivity

8-3-6-1. Definition of concordance, sensitivity, and specificity

The concordance, sensitivity, and specificity were estimated as the indices of predictivity. These indices were estimated using the frequency results obtained from the 2 × 2 contingency table for leukocyte toxicity. The definitions of these indices are summarized

in Table 11 below. This calculation was based on the results decided by the majority for the between-laboratory results for each chemical.

Table 11. Definition of concordance, sensitivity, and specificity

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

Sensitivity = 100 x a/(a+c) or the ability to predict a positive result
Specificity = 100 x d/(b+d) or the ability to predict a negative result
Accuracy = 100 x (a+d)/N

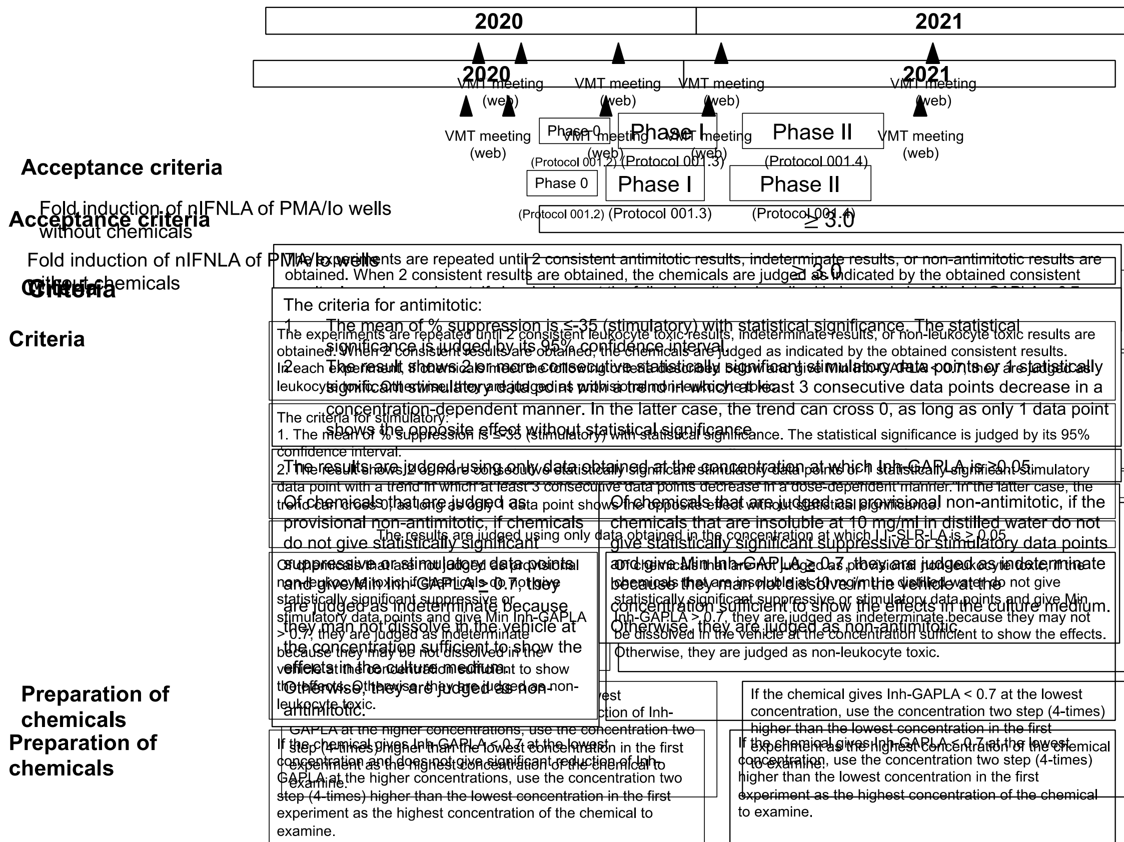
8-4. Quality assurance

Assays and quality assurance were carried out in the spirit of GLP, as 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 were then reviewed by the VMT quality assurance team. The results accurately reflected the raw data.

9. Results

We conducted Phase I and II trials in this validation study. The assay procedure and criteria used to judge antimitotic in the validation trials are summarized in Fig. 7.

Fig. 7. The progress of the validation study for the IL-2 Luc LTT.



9-1. Final criteria for Phase I trial

9-1-1. Acceptance criteria

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

- For each set of experiments, a control experiment examining nIFNLA of 2H4 cells treated with PMA/Io and nIFNLA of non-treated 2H4 cells must be conducted. Then, the fold induction of nIFNLA of 2H4 cells treated with PMA/Ionomycin to nIFNLA 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. Criteria

The experiments are repeated until 2 consistent antimitotic results, indeterminate results, or non-antimitotic results are obtained. When 2 consistent results are obtained, the chemicals are judged as indicated by the obtained consistent results.

In each experiment, if chemicals meet the following criteria described below and have Min Inh-GAPLA < 0.7, they are judged as antimitotic. Otherwise, they are judged as provisional non-antimitotic.

The criteria for stimulatory:

1. The mean of % suppression is ≤-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 stimulatory data points or 1 statistically significant stimulatory data point with a trend in which at least 3 consecutive data points decrease in a concentration-dependent manner. In the latter case, the trend can cross 0, if 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.

Of chemicals that are judged as provisional non-antimitotic, if chemicals do not demonstrate statistically significant suppressive or stimulatory data points, show Min Inh-GAPLA ≥ 0.7, and they are insoluble at 10 mg/mL in distilled water, they are judged as indeterminate because they may be not dissolved in the vehicle at the concentration

sufficient to show the effects in the culture medium. Otherwise, they are judged as non-antimitotic.

9-1-3. Predictivity

The IL-2 Luc LTT is aimed to detect antimitotic effects of chemicals. We first examined the performance of the IL-2 Luc LTT for detecting antimitotic effects of chemicals. Then, we examined its performance for immunosuppressive effects because it has a potential to detect immunosuppression caused by antimitotic effects. To determine the predictivity of the immunotoxicity test, reference data that aid positive or negative judgments are essential. Since no such reference data are currently available to determine the predictability of the IL-2 Luc LTT, we prepared reference data. To make the reference data for immunosuppression, we referred to the rationale for immunotoxicity classification proposed by Luster et al (Luster, 1998) in which they conducted statistical evaluation of data from a screening battery using a tier approach to detect potential immunotoxic compounds in mice (Luster et al., 1992). According to their rationale, a positive reference chemical would either produce a significant concentration-dependent effect in the immune test or significantly alter two or more immune test results at the highest concentration of the chemical tested. Chemicals were classified based on the results 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, using this rationale, we classified chemicals as described in our previous publication (Kimura et al., 2020).

Briefly, we first surveyed the literature, collected the following eight endpoints regarding each chemical used in the study (Table 12), and generated reference data for their immunosuppressive profiles used in the validation study (Appendix 12). The references of the literature that Appendix 12 referred to were shown in Appendix 13. Then, we classified a chemical as an immunosuppressive chemical if it meets Criterion 1 or classified a chemical as anti-mitotic if it meets Criterion 2 (Table 13). The summarized immunotoxicity information and the classifications of the chemicals are also shown in Appendix 12.

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

Table 12. Immunotoxicological data obtained from literature.

Endpoint	Information
Endpoint 1	Decreased antibody response
Endpoint 2	Myelosuppressive or antimitotic
Endpoint 3	Decreased LPS response *in vivo*, *ex vivo*, or *in vitro*
Endpoint 4	Suppressed DHR
Endpoint 5	Suppressed host resistance
Endpoint 6	The NTP data or Tox 21 data indicate immunotoxicity of chemicals
Endpoint 7	Decreased thymus weight
Endpoint 8	Increased or decreased IL-2 mRNA expression or protein production by T cells in *ex vivo*.
Endpoint 9	Increased or decreased IL-2 mRNA expression or protein production by T cells *in vitro*.

Table 13. Criteria to classify immunotoxic chemicals

Criterion	Definition
Criterion 1 (systemic immunotoxicity)	Satisfy one of Endpoints 1 to 9
Criterion 2 (antimitotic effect)	Satisfy Endpoint 2

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

The preliminary test trial (Phase 0) was performed by the participating laboratories following explicit explanations of "IL-2 Luc leukocyte toxicity test (IL-2 Luc LTT) protocol ver.001.2" by the lead laboratory, Tohoku University. Three laboratories participated in the Phase 0 trial of the IL-2 Luc LTT using the three open-labeled chemicals, bleomycin sulfate, dexamethasone, and 6-thioguanine, and conducted one set (three experiments) for each chemical. The response patterns for the three chemicals were similar among the three laboratories. Based on these results, the VMT judged that technical and protocol transfer of the IL-2 Luc LTT was acceptable.

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

9-3-1. Test conditions

A total of five coded chemicals were evaluated by three experimental sets in the Phase I trial based on "IL-2 Luc leukocyte toxicity test (IL-2 Luc LTT) protocol ver.001.3".

In each experimental set, initially two experiments are conducted. If the results are concordant chemical will be classified accordingly, if discordant an additional experiment is conducted. In case of null experiments, they must be repeated.

Chemicals that satisfied the criteria were judged as positive. Chemicals that provided two positive results were judged as antimitotic.

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

The complete results of the Phase I trial are shown in Table 14 and Appendix 10. The within-laboratory reproducibility was 100.0% (15/15), 100% (15/15), and 100.0% (15/15) in Labs A, B, and C, respectively. The average reproducibility was thus 100.0% (15/15).

9-3-3. Between-laboratory variation assessments in Phase I trial

The between-laboratory reproducibility was also 100.0% (15/15).

Table 14. Results of Phase I trial

Chemical	CAS	Set	Lab A	Lab B	Lab C	Concor-dance
Mycophenolic acid	24280-93-1	1st	Antimitotic	Antimitotic	Antimitotic	1
		2nd	Antimitotic	Antimitotic	Antimitotic	
		3rd	Antimitotic	Antimitotic	Antimitotic	
Indomethacin	53-86-1	1st	Antimitotic	Antimitotic	Antimitotic	1
		2nd	Antimitotic	Antimitotic	Antimitotic	
		3rd	Antimitotic	Antimitotic	Antimitotic	
Cyclosporin A		1st	Non	Non	Non	1

	59865-13-3	2nd	Non	Non	Non	
		3rd	Non	Non	Non	
5-Fluorouracil	51-21-8	1st	Antimitotic	Antimitotic	Antimitotic	1
		2nd	Antimitotic	Antimitotic	Antimitotic	
		3rd	Antimitotic	Antimitotic	Antimitotic	
Mannitol	69-65-8	1st	Non	Non	Non	1
		2nd	Non	Non	Non	
		3rd	Non	Non	Non	
Within-laboratory reproducibility (%)			100 (15/15)	100 (15/15)	100 (15/15)	
			Average 100 (15/15)			
Between-laboratory reproducibility (%) (based on majority)						100 (5/5)

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

9-4-1. Test conditions

The Phase II trial to examine between-laboratory reproducibility and predictivity was conducted using a total of 20 coded chemicals and evaluated using one experiment set based on “IL-2 Luc leukocyte toxicity test (IL-2 Luc LTT) protocol ver.001.4”.

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

The complete results of the Phase II trial are shown in Table 15 and Appendix 11. The between-laboratory reproducibility was 90% (18/20).

Table 15. Results of Phase II trial

Chemical	CAS	Lab A	Lab B	Lab C	Based on majority	Concor-dance
Colchicine	64-86-8	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Tetrabromobisphenol AEP	79-94-7	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Fludarabine	21679-14-1	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Nicotinamide	98-92-0	Non	Non	Non	Non	1
Glycidol	556-52-5	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Perfluorooctanoic Acid	335-67-1	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Citral	5392-40-5	Antimitotic	Antimitotic	Non	Antimitotic	0
Cytarabine EP	147-94-4	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Etoposide	33419-42-0	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Gemcitabine hydrochloride	122111-03-9	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
6-Thioguanine	154-42-7	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Cadmium chloride	10108-64-2	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1

Prednisolone	50-24-8	Antimitotic	Antimitotic	Antimitotic	Antimitotic	1
Azathioprine	446-86-6	Non	Non	Non	Non	1
2-Methyoxyacetic Acid	625-45-6	Non	Non	Non	Non	1
4-Chloro-o-phenylenediamine	95-83-0	Non	Antimitotic	Non	Non	0
Dichloroacetic acid	79-43-6	Non	Non	Non	Non	1
n-Nitrosodimethylamine	62-75-9	Non	Non	Non	Non	1
Methyl carbamate	598-55-0	Non	Non	Non	Non	1
Cyclophosphamide monohydrate	6055-19-2	Non	Non	Non	Non	1
Between-laboratory reproducibility (%)						90 (18/20)

9-5. Quality assurance (Appendix 16)

No accidents occurred during the validation study, and all test facilities returned the MSDSs for the test chemicals to JaCVAM in sealed envelopes 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.

All the records (data sheets and record sheets) from the participating laboratories were checked by Dr. Takashi Omori, Kobe Univ. and JaCVAM. The check lists are available at Appendix 17-1 and 17-2. The record sheets include “Chemical records, Solubility test records, Cell culture records and Testing records”. The records are total more than 300 pages and available at JaCVAM website (http:// http://www.jacvam.jp/validation08-login.html). Testing performed as part of the validation study was carried out in non-GLP laboratories in accordance with the

principles of GLP (OECD, 1998) and necessarily include, without being limited to, the use of protocol and adequate recording of data as well as suitable reporting of results and archival record keeping.

We found several insufficient descriptions in the culture of the cells, the preparation and application of test chemicals and data sheets, because a few record sheets were incomplete. Corrected record sheets should be provided before the completion of the validation study and test developers and an independent organization should review them in advance. The results accurately reflect the raw data and the incomplete descriptions did not influence the results.

The following concerns were noted in the validation trials. The QC team requested to revise the protocol after the validation study.

1) The measurements of luciferase activity were done up to 33ºC in all the laboratories. Although the record sheet required a record of actual temperature, there is no description of experimental temperature in protocol. The QC team requested to describe/recommends describing the measured temperature in the protocol if it affects the test results.
2) The culture period after addition of PMA/ionomycin is 6 hr in the protocol. However, the culture period in a laboratory was done up to 7 hr. The QC team requested to revise/recommends revising the protocol based on scientific data.
3) The experiment's number of positive and negative controls are not fixed in the protocol. The QC team requested to describe/recommends describing that the

control tests should be done at the same time as each experiment in the protocol to demonstrate that each experiment was done adequately.

The QC team carefully checked the other results and judged all data to be within acceptable ranges. The QC team recommended to JaCVAM that the validation study performed with GLP laboratories should be move forward.

9-6. Combined results of Phase I and II trials (for between- and within- laboratory reproducibility)

9-6-1. Test conditions

The within- and between-laboratory reproducibility, and the predictivity of the IL-2 Luc LTT, were evaluated using all the results from the phases I and II trials.

9-6-2. Between-laboratory variation assessments from Phase I and II trials (Table 16)

To further evaluate the between-laboratory reproducibility, the results from the Phase I and II trials were combined. The reproducibility for the combined results was 92.0% (23/25), similar to that of the Phase II trial alone.

Table 16. Combined results of Phase I and II trials

Chemical	CAS	Lab A	Lab B	Lab C	Concord-ance
Phase I					
Mycophenolic acid	24280-93-1	Antimitotic/ Antimitotic/ Antimitotic	Antimitotic/ Antimitotic/ Antimitotic	Antimitotic/ Antimitotic/ Antimitotic	1
Indomethacin	53-86-1	Antimitotic/ Antimitotic/ Antimitotic	Antimitotic/ Antimitotic/ Antimitotic	Antimitotic/ Antimitotic/ Antimitotic	1
Cyclosporin A	59865-13-3	Non/Non/Non	Non/Non/Non	Non/Non/Non	1
5-Fluorouracil	51-21-8	Antimitotic/ Antimitotic/ Antimitotic	Antimitotic/ Antimitotic/ Antimitotic	Antimitotic/ Antimitotic/ Antimitotic	1
Mannitol	69-65-8	Non/Non/Non	Non/Non/Non	Non/Non/Non	1
Phase II					
Colchicine	64-86-8	Antimitotic	Antimitotic	Antimitotic	1
Tetrabromobisphenol A EP	79-94-7	Antimitotic	Antimitotic	Antimitotic	1

Fludarabine	21679-14-1	Antimitotic	Antimitotic	Antimitotic	1
Nicotinamide	98-92-0	Non	Non	Non	1
Glycidol	556-52-5	Antimitotic	Antimitotic	Antimitotic	1
Perfluorooctanoic Acid	335-67-1	Antimitotic	Antimitotic	Antimitotic	1
Citral	5392-40-5	Antimitotic	Antimitotic	Non	0
Cytarabine EP	147-94-4	Antimitotic	Antimitotic	Antimitotic	1
Etoposide	33419-42-0	Antimitotic	Antimitotic	Antimitotic	1
Gemcitabine hydrochloride	122111-03-9	Antimitotic	Antimitotic	Antimitotic	1
6-Thioguanine	154-42-7	Antimitotic	Antimitotic	Antimitotic	1
Cadmium chloride	10108-64-2	Antimitotic	Antimitotic	Antimitotic	1
Prednisolone	50-24-8	Antimitotic	Antimitotic	Antimitotic	1
Azathioprine	446-86-6	Non	Non	Non	1
2-Methyoxyacetic Acid	625-45-6	Non	Non	Non	1
4-Chloro-o-phenylenediamine	95-83-0	Non	Antimitotic	Non	0
Dichloroacetic acid	79-43-6	Non	Non	Non	1

n-Nitrosodimethylamine	62-75-9	Non	Non	Non	1
Methyl carbamate	598-55-0	Non	Non	Non	1
Cyclophosphamide monohydrate	6055-19-2	Non	Non	Non	1
Within-laboratory reproducibility (%)		100.0 (5/5)	100.0 (5/5)	100.0 (5/5)	
		Average 100.0 (15/15)			
Between-laboratory reproducibility (%) (Based on majority for Phase I)					92.0 (23/25)

9-7. The predictivity of the IL-2 Luc LTT in the Phase I and Phase II trials

When the predictivity of the IL-2 Luc LTT for immunosuppressive effects was determined (Table 17), the IL-2 Luc LTT performance was 71.4 % (15/21) for mean sensitivity, 100.0 % (4/4) for mean specificity, and 76.0 % (19/25) for mean prediction for the combined data from the Phase I and Phase II trials (Table 18). On the other hand, when the predictivity of the IL-2 Luc LTT for antimitotic effects was determined (Table 5), the IL-2 Luc LTT performance was 76.5 % (13/17) for mean sensitivity, 75.0 % (6/8) for mean specificity, and 76.0 % (19/25) for mean prediction for the combined data from the Phase 1 and Phase 2 trials (Table 18).

Table 17. The judgment of each laboratory in the Phase I and Phase II trials

Chemical	CAS	Classification of chemicals		Judgment by Lab A		Judgment by Lab B		Judgment by Lab C		Judgment by the majority*	
		Immuno-suppressive	Antimitotic	Lab.A/immuno-suppressive	Lab.A/antimitotic	Lab.B/immuno-suppressive	Lab.B/antimitotic	Lab.C/immuno-suppressive	Lab.C/antimitotic	Majority/immuno-suppressive	Majority/antimitotic
Phase I											
Mycophenolic acid	24280-93-1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indomethacin	53-86-1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cyclosporin A	59865-13-3	Yes	No	No	No	No	No	No	No	No	No
5-Fluorouracil	51-21-8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mannitol	69-65-8	No	No	No	No	No	No	No	No	No	No
Phase II											
Colchicine	64-86-8	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tetrabromobisphenol A EP	79-94-7	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fludarabine	21679-14-1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nicotinamide	98-92-0	No	No	No	No	No	No	No	No	No	No
Glycidol	556-52-5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Perfluorooctanoic Acid	335-67-1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Citral	5392-40-5	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Cytarabine EP	147-94-4	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Etoposide	33419-42-0	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gemcitabine hydrochloride	122111-03-9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-Thioguanine	154-42-7	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cadmium chloride	10108-64-2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prednisolone	50-24-8	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Azathioprine	446-86-6	Yes	Yes	No	No	No	No	No	No	No	No
2-Methyoxyacetic Acid	625-45-6	Yes	Yes	No	No	No	No	No	No	No	No
4-Chloro-o-phenylenediamine	95-83-0	No	No	No	No	Yes	Yes	No	No	No	No
Dichloroacetic acid	79-43-6	No	No	No	No	No	No	No	No	No	No
n-Nitrosodimethylamine	62-75-9	Yes	Yes	No	No	No	No	No	No	No	No
Methyl carbamate	598-55-0	Yes	No	No	No	No	No	No	No	No	No
Cyclophosphamide monohydrate	6055-19-2	Yes	Yes	No	No	No	No	No	No	No	No

*: The judgment consistent between the two laboratories.

Table 18. The predictivity of each laboratory in the phase I and Phase II trials

	Judgment by Lab A		Judgment by Lab B		Judgment by Lab C		Judgment by the majority	
Judgment	Immuno-suppressive	Antimitotic	Immuno-suppressive	Antimitotic	Immuno-suppressive	Antimitotic	Immuno-suppressive	Antimitotic
Correct positive	15	13	15	13	14	13	15	13
False negative	6	4	6	4	7	4	6	4
False positive	0	2	1	3	0	1	0	2
Correct negative	4	6	3	5	4	7	4	6
Total	25	25	25	25	25	25	25	25
Performance								
Sensitivity	0.714	0.765	0.714	0.765	0.667	0.765	0.714	0.765
Specificity	1.000	0.750	0.750	0.625	1.000	0.875	1.000	0.750
Accuracy	0.760	0.760	0.720	0.720	0.720	0.800	0.760	0.760

10. Discussion

10-1. Reliability

The IL-2 Luc LTT is based on the modulation of PMA/Io-induced luciferase activity in the IL-2 reporter cell line, 2H4. Therefore, it is important that 2H4 cells retain their ability to induce luciferase activity after PMA/Io stimulation after passage for a sufficient number of times to perform the assay in the long term. We confirmed that a frozen stock of 2H4 cells can be cultured without losing luciferase activity for at least 16 weeks or 33 passages.

Culturing 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, chemicals at graded concentrations are added to the wells of a 96-well culture plate. Then, cells adjusted to the optimum concentration are seeded into each well. After 24-hour incubation, 100 μL of pre-warmed Tripluc Luciferase Assay Reagent 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 LTT is a test method that can significantly reduce human error.

Moreover, the IL-2 Luc LTT 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 promoter activity and cell viability is obtained in each experiment. Furthermore, a single experiment takes only 31 hours, including the time required for chemical preparation and cell plating.

10-2. Between- and within-laboratory reproducibility

We examined within-laboratory reproducibility in the Phase I trial. Lab A, Lab B, and Lab C demonstrated 100%, 100%, and 100% reproducibility, respectively. On the other hand, the between-laboratory reproducibility result for Lab A, Lab B, and Lab C was 92.0 % for the combined data of the Phase I and Phase II trials. These results satisfied the acceptance criteria for the validation study of a within-laboratory reproducibility of at least 80% and a between-laboratory reproducibility of at least 80%.

10-3. Predictivity

10-3-1. The predictivity of the IL-2 Luc LTT in the validation trials

In the combined data from the Phase 1 and Phase 2 trials, the IL-2 Luc LTT performance for immunosuppressive effects was 71.4 % (15/21) for mean sensitivity, 100.0 % (4/4) for mean specificity, and 76.0 % (19/25) for mean prediction. On the other hand, the IL-2 Luc LTT performance for antimitotic effects was 76.5 % (13/17) for mean sensitivity, 75.0 % (6/8) for mean specificity, and 76.0 % (19/25) for mean prediction.

10-3-2. IL-2 Luc LTT judgment for 83 data set chemicals

To clarify the characteristics of the IL-2 Luc LTT, the lead laboratory assessed a total of 85 test compounds including 46 pharmaceutical drugs that were composed of 12 anti-cancer drugs, 8 immunosuppressive drugs with antimitotic effects, 8

immunosuppressive drugs without antimitotic effects, 3 JAK inhibitors, and 15 non-immunosuppressive drugs; and 37 non-pharmaceutical chemicals (23 immunosuppressive and 14 non-immunosuppressive). Seven of the 23 immunosuppressive chemicals have antimitotic properties. (Kimura et al., 2021; Kimura et al., 2020). To determine the performance of the IL-2 Luc LTT for the 37 non-pharmaceutical chemicals, we referred to reference data generated by collecting literature information on the nine endpoints for each chemical (Appendix 14) and identified chemicals that satisfy Criteria 1 (for immunosuppressive effects) or 2 (for antimitotic effects). The summarized immunotoxicity information, together with the classification of each non-pharmaceutical chemicals, are shown in Appendix 14. The judgment of 46 pharmaceutical drugs and 37 non-pharmaceutical chemicals by the IL-2 Luc LTT for antimitotic effects or for immunosuppressive effects is shown in Table 19 and Table 20, respectively. When pharmaceutical drugs were examined, the performance of the IL-2 Luc LTT for antimitotic effects was 68.4 % (13/19) for sensitivity, 91.7 % (22/24) for specificity, and 81.4 % (35/43) for predictivity (Table 24). When non-pharmaceutical chemicals were examined, the performance of the IL-2 Luc LTT for antimitotic effects was 50.0 % (4/8) for sensitivity, 88.9 % (24/27) for specificity, and 80.0 % (28/35) for predictivity (Table 25). When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the IL-2 Luc LTT for antimitotic effects was 63.0 % (17/27) for sensitivity, 90.2 % (46/51) for specificity, and 80.8 % (63/78) for predictivity. When the applicability domain was taken into consideration, the performance of the IL-2 Luc LTT for antimitotic effects

was 62.5 % (15/24) for sensitivity, 90.2 % (46/51) for specificity, and 81.3 % (61/75) for predictivity (Table 26).

On the other hand, when pharmaceutical drugs were examined, the performance of the IL-2 Luc LTT for immunosuppressive effects was 43.3 % (13/30) for sensitivity, 84.6 % (11/13) for specificity, and 55.8 % (24/43) for predictivity (Table 25). When non-pharmaceutical chemicals were examined, the performance of the IL-2 Luc LTT for immunosuppressive effects was 26.1 % (6/23) for sensitivity, 91.7 % (11/12) for specificity, and 48.6 % (17/35) for predictivity (Table 25). When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the IL-2 Luc LTT for immunosuppressive effects was 35.8 % (19/53) for sensitivity, 88.0 % (22/25) for specificity, and 52.6 % (41/78) for predictivity. After considering the applicability domain, the performance of the IL-2 Luc LTT for immunosuppressive effects was 34.0 % (17/50) for sensitivity, 88.0 % (22/25) for specificity, and 52.0 % (39/75) for predictivity. (Table 26).

10-3-3. Application of the IL-2 Luc LTT with IL-2 Luc as part of an IATA approach

To examine whether the combination of the IL-2 Luc LTT and the IL-2 Luc assay improve the performance of the IL-2 Luc LTT for immunosuppressive effects, we first examined the performance of the IL-2 Luc assay alone (Table 21 and 23). The IL-2 Luc assay that we have reported previously (Arch Toxicol 92: 2043-2054, 2018: Toxicol in Vitro 66: 104832, 2020) and are currently under the validation is composed of three

different endpoints: suppression, stimulation, and no effect. Since the IL-2 Luc assay is to detect chemicals targeting T cell function, regardless of whether they are inhibitory or stimulatory, stimulation was included as a potential judgement. In contrast, since the IL-2 Luc assay used in the combination with the IL-2 Luc LTT was aimed to detect chemicals that induce immunosuppression, we considered stimulation judgment as no effect. In Table 23, we presented the highest peak plasma concentrations (Cmax) of each pharmaceutical drug and its lowest observed effect level (LOEL) values obtained from the IL-2 Luc assay and the IL-2 Luc LTT. When pharmaceutical drugs were examined, the performance of the IL-2 Luc assay for immunosuppressive effects was 48.4 % (15/31) for sensitivity, 66.7 % (10/15) for specificity, and 54.3 % (25/46) for predictivity (Table 24). When non-pharmaceutical chemicals were examined, the performance of the IL-2 Luc assay for immunosuppressive effects was 60.9 % (14/23) for sensitivity, 64.3 % (9/14) for specificity, and 62.2 % (23/37) for predictivity (Table 25). When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the IL-2 Luc assay for immunosuppressive effects was 53.7 % (29/54) for sensitivity, 59.0 % (19/29) for specificity, and 57.8 % (48/83) for predictivity (Table 26).

Next, we combined these 2 assays based on the approach that drugs are judged as immunosuppressive when at least one of these 2 assays gives positive results. When pharmaceutical drugs were examined, the performance of the combined assay for immunosuppressive effects was 90.0 % (27/30) for sensitivity, 61.5 % (8/13) for specificity, and 76.1% (33/43) for predictivity (Table 24). When non-pharmaceutical

chemicals were examined, the performance of the combined assay for immunosuppressive effects was 82.6 % (19/23) for sensitivity, 58.3 % (7/12) for specificity, and 74.3 % (26/35) for predictivity (Table 25). When pharmaceutical drugs and non-pharmaceutical chemicals were examined together, the performance of the combined assay for immunosuppressive effects was 86.8 % (46/53) for sensitivity, 60.0 % (15/25) for specificity, and 78.2 % (61/78) for predictivity (Table 26). After considering the applicability domain, the performance of the combined assay was 88.8 % (44/50) for sensitivity, 60.0 % (15/25) for specificity, and 78.7 % (59/75) for predictivity.

Table 19. IL-2 Luc LTT results and concordance with classification of 46 pharmaceutical drugs

No.	Pharmaceutical drugs	Classification		IL-2 Luc LTT Results and Judgment			
		Antimitotic	Immuno-suppressive	Min Inh-GAPLA<0.7	Min %suppression<-35	Antimitotic	Immuno-suppressive
1	Adriamycin/Doxorubicin	Y	Y	Y	Y	Y	Y
2	Bleomycin	Y	Y	Y	Y	Y	Y
3	Busulfan	Y	Y	N	N	I	I
4	Carmustine (BCNU)	Y	Y	Y	N	N	N
5	Cytosar-U/Cytarabine	Y	Y	Y	Y	Y	Y
6	Etoposide	Y	Y	Y	Y	Y	Y
7	Fludarabine	Y	Y	Y	Y	Y	Y
8	5-Fluorouracil	Y	Y	Y	Y	Y	Y
9	Gemcitabine	Y	Y	Y	Y	Y	Y
10	Paclitaxel	Y	Y	Y	N	N	N
11	Teniposide	Y	Y	Y	Y	Y	Y
12	6-Thioguanine	Y	Y	Y	Y	Y	Y
13	Azathioprine	Y	Y	Y	N	N	N
14	6-Mercaptoprine	Y	Y	Y	N	N	N
15	Leflunomide	Y	Y	N	N	N	N
16	Methotrexate	Y	Y	N	Y	N	N
17	Mizoribine	Y	Y	Y	Y	Y	Y
18	Mycophenolic acid	Y	Y	Y	Y	Y	Y
19	Rapamycin	Y	Y	Y	Y	Y	Y
20	Dexamethasone	N	Y	Y	N	N	N
21	Cyclosporine	N	Y	N	N	N	N
22	Tacrolimus	N	Y	N	N	N	N
23	Iguratimod	N	Y	N	N	N	N
24	Colchicine	Y	Y	Y	Y	Y	Y
25	Sulfasalazine	N	Y	N	N	N	N
26	Minocycline	N	Y	Y	N	N	N
27	Apremilast	N	Y	N	Y	N	N
28	Chloroquine	N	Y	Y	N	N	N
29	Baricitinib	N	Y	N	N	N	N
30	Ruxolitinib	N	Y	N	N	N	N
31	Tofacitinib	N	Y	N	N	N	N
32	Acetoaminophen	N	N	Y	Y	Y	Y
33	Indomethacine	N	N	Y	Y	Y	Y
34	Pirfenidone	N	N	N	N	I	I
35	Digoxin	N	N	Y	N	N	N
36	Warfarin	N	N	N	N	N	N
37	Li2Co3	N	N	N	N	N	N
38	Acyclovir	N	N	N	N	I	I
39	Retrovir	N	N	Y	N	N	N
40	Isoniazide	N	N	N	N	N	N
41	Terbinefine	N	N	Y	N	N	N
42	Diphenhydramine	N	N	Y	N	N	N
43	Chlorothiazide	N	N	N	Y	N	N
44	Pravastatin	N	N	N	N	N	N
45	Lansoprazole	N	N	Y	N	N	N
46	Mannitol	N	N	N	N	N	N

Y = yes, N= no, I= indeterminate

Color designation for assay judgements compared to literature classification (criteria 1): green = correct, red = false

Table 20. IL-2 Luc LTT results and concordance with literature-based classification for non-pharmaceutical chemicals

No.	Chemicals	Classification		IL-2 Luc LTT Results and Judgment			
		Antimitotic	Immuno-suppressive	Min Inh-GAPLA<0.7	Min %suppression <-35	Antimitotic	Immuno-suppressive
47	Dibenzopyrene	Y	Y	N	N	N	N
48	Mercuric chloride	Y	Y	N	N	N	N
49	Diesel exhaust particles	Y	Y	N	N	N	N
50	2-Methyoxyacetic Acid	Y	Y	N	N	N	N
51	Cadmium chloride	Y	Y	Y	Y	Y	Y
52	Glycidol	Y	Y	Y	Y	Y	Y
53	Perfluorooctanoic Acid	Y	Y	Y	Y	Y	Y
54	Tetrabromobisphenol A EP	Y	Y	Y	Y	Y	Y
55	2-Mercaptobenzothiazole	N	Y	Y	Y	Y	Y
56	2,4-Diaminotoluene	N	Y	N	N	N	N
57	Aluminum chloride	N	Y	N	N	N	N
58	Citral	N	Y	Y	Y	Y	Y
59	Cobalt chloride	N	Y	N	N	N	N
60	Diethanolamin	N	Y	N	N	N	N
61	Formaldehyde	N	Y	N	N	N	N
62	Hydrogen peroxide	N	Y	N	N	N	N
63	Isophorone diisocyanate	N	Y	N	N	N	N
64	Methanol	N	Y	N	N	N	N
65	Nickel sulfate	N	Y	N	N	N	N
66	Sodium dodecyl sulfate	N	Y	N	N	N	N
67	Dibutyl phthalate	N	Y	N	N	N	N
68	Lead(II) acetate	N	Y	N	N	N	N
69	Pyrimethamine	N	Y	N	N	N	N
70	Benzethonium chloride	N	N	N	N	N	N
71	Dimethyl sulfoxide	N	N	N	N	N	N
72	Ethanol	N	N	N	N	N	N
73	Hexachlorobenzene	N	N	I	I	I	I
74	Histamine	N	N	N	N	N	N
75	Magnesium sulfate	N	N	N	N	N	N
76	Nitrofurazone	N	N	N	N	N	N
77	*p*-Nitroaniline	N	N	N	N	N	N
78	Pentamidine isethionate	N	N	N	N	N	N
79	Trichloroethylene	N	N	I	I	I	I
80	Chloroplatinic acid	N	N	Y	Y	Y	Y
81	Sodium bromate	N	N	N	N	N	N
82	Triethanolamine	N	N	N	N	N	N
83	Dichloroacetic acid	N	N	N	N	N	N

Y = yes, N = no, I= indeterminate

Color designation for assay judgements compared to literature classification (criteria 1): green = correct, red = false

Table 21. Combination of IL-2 Luc and IL-2 Luc LTT judgements and concordance with classification of 46 pharmaceutical drugs.

No.	Pharmaceutical drugs	Classification		IL-2 Luc LTT Judgment		IL-2 Luc assay Judgment	Combined	Combined + AD*
		Antimitotic	Immuno-suppressive	Antimitotic	Immuno-suppressive	Immuno-suppressive	Judge	Judge
1	Adriamycin/Doxorubicin	Y	Y	Y	Y	Y	Y	Y
2	Bleomycin	Y	Y	Y	Y	N	Y	Y
3	Busulfan	Y	Y	I	I	N	I	I
4	Carmustine (BCNU)	Y	Y	N	N	Y	Y	Y
5	Cytosar-U/Cytarabine	Y	Y	Y	Y	N	Y	Y
6	Etoposide	Y	Y	Y	Y	N	Y	Y
7	Fludarabine	Y	Y	Y	Y	N	Y	Out of AD
8	5-Fluorouracil	Y	Y	Y	Y	N	Y	Y
9	Gemcitabine	Y	Y	Y	Y	N	Y	Y
10	Paclitaxel	Y	Y	N	N	Y	Y	Y
11	Teniposide	Y	Y	Y	Y	N	Y	Y
12	6-Thioguanine	Y	Y	Y	Y	N	Y	Y
13	Azathioprine	Y	Y	N	N	Y	Y	Out of AD
14	6-Mercaptopurine	Y	Y	N	N	N	N	Out of AD
15	Leflunomide	Y	Y	N	N	N	Y	Y
16	Methotrexate	Y	Y	N	N	N	N	N
17	Mizoribine	Y	Y	Y	Y	N	Y	Y
18	Mycophenolic acid	Y	Y	Y	Y	N	Y	Y
19	Rapamycin	Y	Y	Y	Y	N	Y	Y
20	Dexamethasone	N	Y	N	N	Y	Y	Y
21	Cyclosporine	N	Y	N	N	Y	Y	Y
22	Tacrolimus	N	Y	N	N	Y	Y	Y
23	Iguratimod	N	Y	N	N	Y	Y	Y
24	Colchicine	Y	Y	Y	Y	Y	Y	Y
25	Sulfasalazine	N	Y	N	N	Y	Y	Y
26	Minocycline	N	Y	N	N	Y	Y	Y
27	Apremilast	N	Y	N	N	N	N	N
28	Chloroquine	N	Y	N	N	Y	Y	Y
29	Baricitinib	N	Y	N	N	Y	Y	Y
30	Ruxolitinib	N	Y	N	N	Y	Y	Y
31	Tofacitinib	N	Y	N	N	Y	Y	Y
32	Acetoaminophen	N	N	Y	Y	N	Y	Y
33	Indomethacine	N	N	Y	Y	Y	Y	Y
34	Pirfenidone	N	N	I	I	Y	I	I
35	Digoxin	N	N	N	N	Y	Y	Y
36	Warfarin	N	N	N	N	N	N	N
37	Li2Co3	N	N	N	N	N	N	N
38	Acyclovir	N	N	I	I	N	I	I
39	Retrovir	N	N	N	N	N	N	N
40	Isoniazide	N	N	N	N	N	N	N
41	Terbinefine	N	N	N	N	N	N	N
42	Diphenhydramine	N	N	N	N	Y	Y	Y
43	Chlorothiazide	N	N	N	N	N	N	N
44	Pravastatin	N	N	N	N	N	N	N
45	Lansoprazole	N	N	N	N	Y	Y	Y
46	Mannitol	N	N	N	N	N	N	N

Y = yes, N= no, I= indeterminate

Color designation for assay judgements compared to literature classification (criteria 1): green = correct, red = false

AD*: take applicability domain in consideration

Table 22. Comparison of in vivo C_{max} and LOEL values from the IL-2 Luc and IL-2 Luc LTT assays

No.	Pharmaceutical drugs	Classification		IL-2 Luc assay		IL-2 Luc LTT		
		Immuno-suppressive	Cmax in humans (μg/ml)	Judgment	Average LOEL (μg/mL)	Judgment	Min Inh-GAPLA<0.7 Average LOEL (μg/mL)	Min %suppression<-35 Average LOEL (μg/mL)
1	Adriamycin/Doxorubicin	Yes	3.66	Y	1.56	Y	0.15	0.05
2	Bleomycin	Yes	4.00	N		Y	50.00	4.69
3	Busulfan	Yes	1.22	N		I		
4	Carmustine (BCNU)	Yes	4.15	Y	12.50	N	14.58	
5	Cytosar-U/Cytarabine	Yes	13.20	N		Y	31.31	1.96
6	Etoposide	Yes	19.70	N		Y	12.50	0.39
7	Fludarabine	Yes	1.13	N	1.95	Y	250.00	0.49
8	5-Fluorouracil	Yes	48.40	N		Y	0.98	0.98
9	Gemcitabine	Yes	23.50	N		Y	1.10	0.98
10	Paclitaxel	Yes	3.35	Y	6.25	N	0.01	
11	Teniposide	Yes	15.20	N		Y	1.17	0.20
12	6-Thioguanine	Yes	0.01	N		Y	0.37	0.24
13	Azathioprine	Yes	0.07	Y	58.48	N	20.78	
14	6-Mercaptoprine	Yes	0.07	N		N	25.00	
15	Leflunomide	Yes	10.10	N	3.13	N		
16	Methotrexate	Yes	0.59	N		N		0.39
17	Mizoribine	Yes	9.60	N		Y	39.00	6.50
18	Mycophenolic acid	Yes	34.00	N	0.40	Y	0.14	0.12
19	Rapamycin	Yes	0.00	N		Y	0.00	0.01
20	Dexamethasone	Yes	0.09	Y	31.10	N	125.00	
21	Cyclosporine	Yes	0.98	Y	0.00	N		
22	Tacrolimus	Yes	0.04	Y	0.00	N		
23	Iguratimod	Yes	1.17	Y	15.63	N		
24	Colchicine	Yes	0.01	Y	0.27	Y	2.15	2.15
25	Sulfasalazine	Yes	15.60	Y	45.50	N		
26	Minocycline	Yes	4.80	Y	18.52	N	6.45	
27	Apremilast	Yes	0.34	N	4.20	N		4.25
28	Chloroquine	Yes	0.94	Y	4.15	N	15.63	
29	Baricitinib	Yes	0.15	Y	1.25	N		
30	Ruxolitinib	Yes	1.65	Y	2.00	N		
31	Tofacitinib	Yes	0.14	Y	2.50	N		
32	Acetoaminophen	No	9.40	N	100.00	Y	500.00	23.44
33	Indomethacine	No	1.32	Y	93.75	Y	56.25	40.63
34	Pirfenidone	No	10.60	Y	250.00	I		
35	Digoxin	No	0.00	Y	0.07	N	0.05	
36	Warfarin	No	0.69	N		N		
37	Li2Co3	No	110.00	N		N		
38	Acyclovir	No	0.94	N		I		
39	Retrovir	No	0.55	N	75.00	N	333.33	
40	Isoniazide	No	6.03	N		N		
41	Terbinefine	No	0.73	N		N	25.00	
42	Diphenhydramine	No	0.08	Y	11.72	N	37.50	
43	Chlorothiazide	No	0.22	N		N		75.00
44	Pravastatin	No	0.03	N		N		
45	Lansoprazole	No	0.53	Y	15.63	N	18.75	
46	Mannitol	No	12200.00	N		N		

Y = yes, N= no

Color designation for assay judgements compared to literature classification (criteria 1): green = correct, red = false

Table 23. Combination of IL-2 Luc and IL-2 Luc LTT judgements and concordance with literature-based classification for non-pharmaceutical chemicals

No.	Chemicals	Classification		IL-2 Luc LTT Judgment		IL-2 Luc assay Judgmenty	Combined
		Antimitotic	Immuno-suppressive	Antimitotic	Immuno-suppressive	Immuno-suppressive	Immuno-suppressive
47	Dibenzopyrene	Y	Y	N	N	N	N
48	Mercuric chloride	Y	Y	N	N	Y	Y
49	Diesel exhaust particles	Y	Y	N	N	Y	Y
50	2-Methyoxyacetic Acid	Y	Y	N	N	N	N
51	Cadmium chloride	Y	Y	Y	Y	N	Y
52	Glycidol	Y	Y	Y	Y	N	Y
53	Perfluorooctanoic Acid	Y	Y	Y	Y	N	Y
54	Tetrabromobisphenol A EP	Y	Y	Y	Y	N	Y
55	2-Mercaptobenzothiazole	N	Y	Y	Y	N	Y
56	2,4-Diaminotoluene	N	Y	N	N	N	N
57	Aluminum chloride	N	Y	N	N	Y	Y
58	Citral	N	Y	Y	Y	Y	Y
59	Cobalt chloride	N	Y	N	N	Y	Y
60	Diethanolamin	N	Y	N	N	Y	Y
61	Formaldehyde	N	Y	N	N	Y	Y
62	Hydrogen peroxide	N	Y	N	N	Y	Y
63	Isophorone diisocyanate	N	Y	N	N	Y	Y
64	Methanol	N	Y	N	N	N	N
65	Nickel sulfate	N	Y	N	N	Y	Y
66	Sodium dodecyl sulfate	N	Y	N	N	Y	Y
67	Dibutyl phthalate	N	Y	N	N	Y	Y
68	Lead(II) acetate	N	Y	N	N	Y	Y
69	Pyrimethamine	N	Y	N	N	Y	Y
70	Benzethonium chloride	N	N	N	N	Y	Y
71	Dimethyl sulfoxide	N	N	N	N	N	N
72	Ethanol	N	N	N	N	N	N
73	Hexachlorobenzene	N	N	I	I	N	I
74	Histamine	N	N	N	N	N	N
75	Magnesium sulfate	N	N	N	N	N	N
76	Nitrofurazone	N	N	N	N	Y	Y
77	*p*-Nitroaniline	N	N	N	N	Y	Y
78	Pentamidine isethionate	N	N	N	N	Y	Y
79	Trichloroethylene	N	N	I	I	N	I
80	Chloroplatinic acid	N	N	Y	Y	Y	Y
81	Sodium bromate	N	N	N	N	N	N
82	Triethanolamine	N	N	N	N	N	N
83	Dichloroacetic acid	N	N	N	N	N	N

Y = yes, N= no, I= indeterminate
Color designation for assay judgements compared to literature classification (criteria 1): green = correct, red = false

Table 24. The performance of the IL-2 Luc LTT, the IL-2 Luc assay, and the combination when pharmaceutical drugs were examined

Pharmaceutical drugs	IL-2 Luc assay	IL-2 Luc LTT		Combined	Combined + AD*
	Immuno-suppressive	Antimitotic	Immuno-suppressive	Judge	
Correct positive	15	13	13	27	25
False negative	16	6	17	3	2
False positive	5	2	2	5	5
Correct negative	10	22	11	8	8
Indeterminate		3	3	3	3
Total	46	46	46	46	43
Sensitivity	0.484	0.684	0.433	0.900	0.926
Specificity	0.667	0.917	0.846	0.615	0.615
Accuracy	0.543	0.814	0.558	0.814	0.825

AD*: take applicability domain in consideration

Table 25. The performance of the IL-2 Luc LTT, the IL-2 Luc assay, and the combination when non-pharmaceutical chemicals were examined

Non-pharmaceutical chemicals	IL-2 Luc assay	IL-2 Luc LTT		Combined
	Immuno-suppressive	Antimitotic	Immuno-suppressive	Judge
Correct positive	14	4	6	19
False negative	9	4	17	4
False positive	5	3	1	5
Correct negative	9	24	11	7
Indeterminate		2	2	2
Total	37	37	37	37
Sensitivity	0.609	0.500	0.261	0.826
Specificity	0.643	0.889	0.917	0.583
Accuracy	0.622	0.800	0.486	0.743

Table 26. The performance of the IL-2 Luc LTT, the IL-2 Luc assay, and the combination when pharmaceutical drugs and non-pharmaceutical chemicals were examined

Drugs and chemicals	IL-2 Luc assay	IL-2 Luc LTT		Combined	Combined + AD*
	Immuno-suppressive	Antimitotic	Immuno-suppressive	Judge	Judge
Correct positive	29	17	19	46	44
False negative	25	10	34	7	6
False positive	10	5	3	10	10
Correct negative	19	46	22	15	15
Indeterminate		5	5	5	5
Total	83	83	83	83	80
Sensitivity	0.537	0.630	0.358	0.868	0.880
Specificity	0.655	0.902	0.880	0.600	0.600
Accuracy	0.578	0.808	0.526	0.782	0.787

AD*: take applicability domain in consideration

10-4. The possible reason for the false negative judgment for carmustin, paclitaxel, leflunomide, and methotrexate (MTX) by the IL-2 Luc LTT.

The examination of 37 pharmaceutical drugs by the lead laboratory elucidated three mechanisms for false negative judgments by the IL-2 Luc LTT. The first one is due to significant suppression of the IL-2 Luc assay by drugs with antimitotic effects, such as carmustin, paclitaxel, and leflunomide. Since the criteria of the IL-2 Luc LTT for positive judgment includes “the mean of % suppression is ≤-35 (stimulatory) with statistical significance”, the IL-2 Luc LTT cannot judge drugs that have inhibitory effects on IL-2 transcription as positive which increase “the mean of % suppression is >35 with statistical significance”.

The second reason is due to the requirement of metabolic activation for drugs to demonstrate antimitotic effects. Of the 46 drugs tested, AZ, 6-mercaptoprine, and fludarabine are known to be metabolized to generate an active metabolite (Bradford and

Shih, 2011; Gandhi and Plunkett, 2002). The IL-2 Luc LTT judged AZ and 6-meracaptoprine as non-antimitotic, while 6-thioguanine, the active metabolite of these 2 drugs, was judged as antimitotic. Although fludarabine was judged as positive by the IL-2 Luc LTT, the LOEL for Min Inh-GAPLA<0.7 was 250 μg/mL, which was much higher than the expected plasma concentration of the patients treated with fludarabine.

The third one is due to the mechanism underlying the false judgment of MTX. MTX was also judged as negative because it did not show significant cytotoxicity even at 200 μg/mL. Similar results were reported by Pessina et al. (Pessina et al., 2003) in which they could not determine the IC90 of MTX in either human umbilical cord blood cells or murine bone marrow cells. There may be some underlying mechanisms in *in vitro* culture that protect proliferating cells from MTX.

10-5. Limitation and applicability domain of the IL-2 Luc LTT

The IL-2 Luc LTT is applicable to test chemicals soluble or that form a stable dispersion but shares limitations common to many suspension cell-based assays when testing highly hydrophobic substances. Chemicals that interfere with luciferase or luminescence can confound its activity/measurement (Thorne et al. 2010) and thus these chemicals are out of the applicability domain. In addition, the following limitations should be noted: 1) the use of PMA/Io as a stimulant bypasses signaling through the T cell receptor and the subsequent intracellular signaling events that precede activation of phospholipase C, and therefore precludes detection of chemicals that act on those

upstream signaling molecules (Ohtsuka et al., 1996); (2) the Jurkat T cell line (from which 2H4 cells are derived) are demonstrated to be suitable for examining the molecular mechanism underlying immunotoxicity (Shao et al., 2013), they may lack several key proteins involved in the activation of normal T cells in response to TCR stimulation, and therefore may not be able to detect effects of chemicals that act on those key proteins.

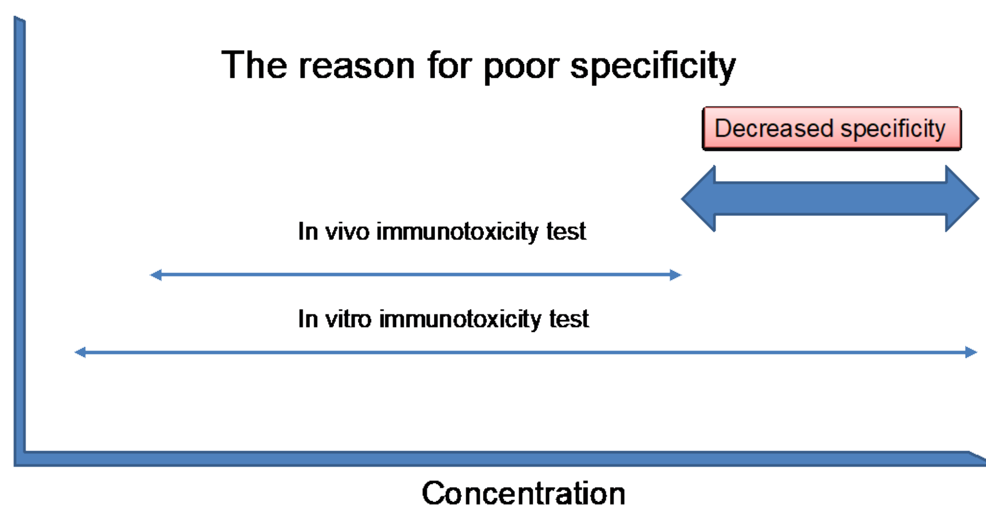
Of the 46 drugs tested, AZ, 6-mercaptoprine, and fludarabine are known to be metabolized to generate the active metabolite (Bradford and Shih, 2011; Gandhi and Plunkett, 2002). The IL-2 Luc LTT judged AZ and 6-meracaptoprine as non-antimitotic, while 6-thioguanine, the active metabolite of these 2 drugs, was judged as antimitotic. Although fludarabine was judged as positive by the IL-2 Luc LTT, the lowest observed effect level (LOEL) for Min Inh-GAPLA<0.7 was 250 μg/mL, which was much higher than the expected plasma concentration of the patients treated with fludarabine. 2H4 cells are not supposed to express the enzymes required to metabolize and activate drugs and thus drugs that need metabolic activation are out of the applicability domain.

10-6. The possible reason for the poor specificity of the IL-2 Luc LTT

Although the sensitivity and the predictivity of the combination assay seemed to be acceptable, the specificity was not satisfactory. In general, there is a marked difference in the concentrations of chemicals used in *in vivo* and *in vitro* tests. Usually, much higher concentrations can be used *in vitro* compared with *in vivo*. Indeed, the IL-2 Luc

LTT and IL-2 Luc assay examined the effects of chemicals starting at the highest concentration of 200 µg/mL. In contrast, the highest peak plasma concentrations (Cmax) of the 46 pharmaceutical drugs using customary dose regimes were <50 µg/mL. In addition, it is unlikely that the plasma concentration of non-pharmaceutical chemicals will exceed that of pharmaceutical drugs in most cases. Therefore, it is likely that false-positive judgments by the IL-2 Luc assay or IL-2 Luc LTT are due to the effects of chemicals at concentrations that have not been examined in *in vivo* experiments, thus resulting in low assay specificity (Fig. 8).

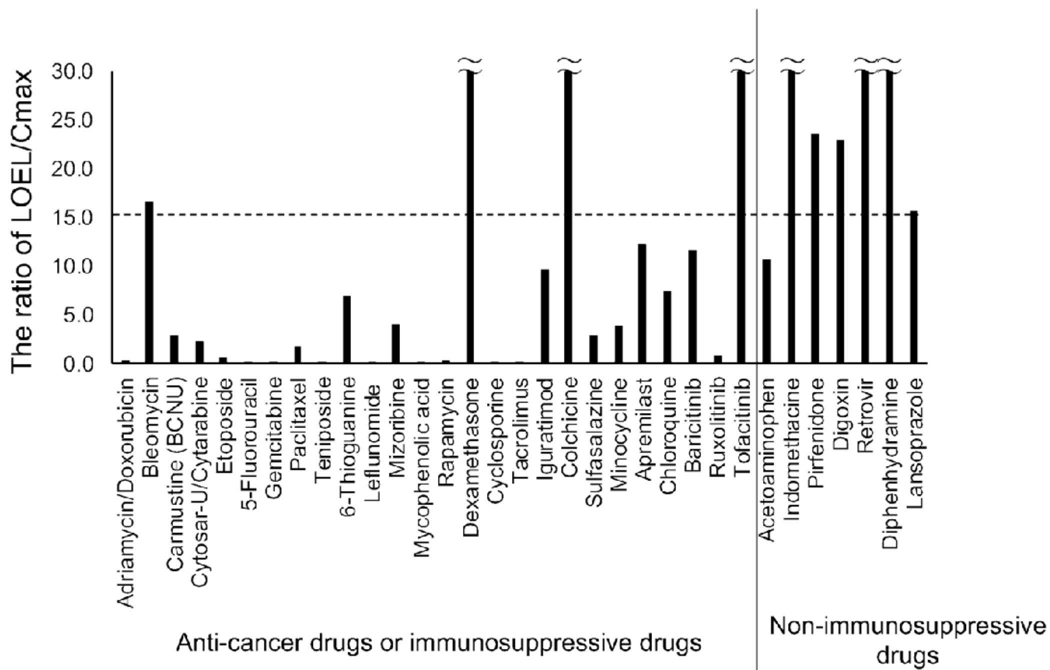
Fig. 8. The possible reason for the poor specificity of the IL-2 Luc LTT



As long as exact LOELs cannot be obtained, the IL-2 Luc assay, the IL-2 Luc LTT, and their combination can be used only for hazard identification. However, although the exact LOELs were not known, the dissociation between the LOELs and Cmax values of anti-cancer drugs or immunosuppressive drugs tends to be much smaller than that of

non-immunosuppressive drugs. We therefore calculated the ratio of LOEL/Cmax of the drugs and found that 22 of the 26 (84.6%) anti-cancer and immunosuppressive drugs and only 1 of the 7 (14.3%) non-immunosuppressive drugs showed a LOEL/Cmax ratio below 15 (Fig. 9). If we tentatively judged the drugs whose LOELs were more than 15 times higher than Cmax as negative, the performance of the combined assay was 88.9 % (24/27) for sensitivity, 92.3 % (12/13) for specificity and 90.0 % (36/40) for accuracy after considering the applicability domain.

Fig. 9. The ratio of LOEL/Cmax for each drug.



The LOEL of each drug was obtained as follows. If drugs are judged as positive by the IL-2 Luc LTT, we selected the higher of the LOEL for Inh-GAPLA<0.7 and the LOEL for % suppression<-35. If drugs gave positive judgment for both the IL-2 Luc assay and the IL-2 Luc LTT, the higher of the LOEL for the IL-2 Luc assay and the LOEL for the IL-2 Luc LTT was selected as the LOEL of the combined assay.

10-7. Regulatory application of the IL-2 Luc LTT

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 due to immunosuppression. 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 (World Health Organization (WHO, 2012). Currently, the assessment of chemical immunotoxicity relies mainly on animal models and assays that characterize immunosuppression and sensitization. However, animal trials have many drawbacks, such as high cost, ethical concerns, and questionable relevance to risk assessment for humans, that they cannot be used to screen the immunotoxicity of more than 130 million chemicals. Therefore, there is an urgent need to develop alternative testing methods and assessment strategies to reduce the use of laboratory animals and, if possible, replace animals used in scientific trials (Adler et al. 2011). To date, however, there are no OECD test guidelines to detect chemical immunotoxicity *in vitro*. We would therefore like to propose the IL-2 Luc LTT in

addition to the IL-2 Luc assay an *in vitro* alternative test methods for immunotoxicity, especially immunosuppression by chemicals.

11. Conclusion

We conducted the validation of the IL-2 Luc LTT and could demonstrate the acceptable within- and between-laboratory reproducibility of this assay. The IL-2 Luc LTT can detect immunosuppressive effects of chemicals caused by the antimitotic effect that cannot be detected by the IL-2 Luc assay. Indeed, the combination of these two assays increases the identification of most immunosuppressive drugs used for organ transplantation, bone marrow transplantation, or treatment of collagen diseases. The performance of the combination of the two assays together for pharmaceutical drugs improves to 95.0% (38/40) after considering the applicability domain. Even for the immunosuppressive effects of non-pharmaceutical chemicals, the performance of the combination of assays is improved to 74.3 % (26/35).

Considering that compounds can target different pathways and immune cells, the assessment of immunotoxicity will require the combination with other assays to further improve the predictivity. The IL-2 Luc LTT is an indispensable component for the IATA for immunotoxicity. We would therefore like to propose the IL-2 Luc LTT as the OECD test guideline for *in vitro* immunotoxicity tests.

12. Acknowledgement

This validation study was supported by Grants-in-Aid from 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 voluntary work by the participating 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

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

Barber, R.D., Harmer, D.W., Coleman, R.A., et al. (2005), GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol Genomics* 21: 389-395, 10.1152/physiolgenomics.00025.2005

Bradford, K., Shih, D.Q. (2011), Optimizing 6-mercaptopurine and azathioprine therapy in the management of inflammatory bowel disease. *World J Gastroenterol* 17: 4166-4173, 10.3748/wjg.v17.i37.4166

Edwards, D.R., Denhardt, D.T. (1985), A study of mitochondrial and nuclear transcription with cloned cDNA probes. Changes in the relative abundance of mitochondrial transcripts after stimulation of quiescent mouse fibroblasts. *Exp Cell Res* 157: 127-143,

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

Gandhi, V., Plunkett, W. (2002), Cellular and clinical pharmacology of fludarabine. *Clin Pharmacokinet* 41: 93-103, 10.2165/00003088-200241020-00002

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 Immunotoxicoll* 2: 61-83, 10.1080/15476910590965832

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. (2015), Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization. *Antimitoticol In Vitro* 29: 1816-1830, 10.1016/j.tiv.2015.07.006

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 Toxicl* 92: 2043-2054, 10.1007/s00204-018-2199-7

Kimura, Y., Terui, H., Fujimura, C., et al. (2021), Optimization of the IL-2 Luc assay for immunosuppressive drugs: a novel in vitro immunotoxicity test with high sensitivity and predictivity. *Arch Toxicol* 95: 2755-2768, 10.1007/s00204-021-03101-4

Kimura, Y., Yasuno, R., Watanabe, M., et al. (2020), An international validation study of the IL-2 Luc assay for evaluating the potential immunotoxic effects of chemicals on T cells and a proposal for reference data for immunotoxic chemicals. *Antimitoticol In Vitro* 66: 104832, 10.1016/j.tiv.2020.104832

Langezaal, I., Hoffmann, S., Hartung, T., et al. (2002), Evaluation and prevalidation of an immunotoxicity test based on human whole-blood cytokine release. *Alternatives to laboratory animals : 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

Luster, A.D. (1998), Chemokines--chemotactic cytokines that mediate inflammation. *New Eng J Med* 338: 436-445,

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

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

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

Mori, R., Wang, Q., Danenberg, K.D., et al. (2008), Both beta-actin and GAPDH are useful reference genes for normalization of quantitative RT-PCR in human FFPE tissue samples of prostate cancer. *Prostate* 68: 1555-1560, 10.1002/pros.20815

Nakajima, Y., Ohmiya, Y. (2010), Bioluminescence assays: multicolor luciferase assay, secreted luciferase assay and imaging luciferase assay. *Expert Opin Drug Discov* 5:

835-849, 10.1517/17460441.2010.506213

OECD 2018. Good Cell Culture Practice (GCCP)", in Guidance Document on Good In Vitro Method Practices (GIVIMP), *OECD Guidance document on Good In vitro Method Practices (GIVIMP)*. OECD publishing. Paris.

Oliveira, J.G., Prados, R.Z., Guedes, A.C., et al. (1999), The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase is inappropriate as internal control in comparative studies between skin tissue and cultured skin fibroblasts using Northern blot analysis. *Arch Dermatol Res* 291: 659-661,

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 Pharamacol* 254: 245-255, 10.1016/j.taap.2011.04.017

Shao, J., Berger, L.F., Hendriksen, P.J., et al. (2014), Transcriptome-based functional classifiers for direct immunotoxicity. *Arch Toxicol* 88: 673-689, 10.1007/s00204-013-1179-1

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

Thellin, O., Zorzi, W., Lakaye, B., et al. (1999), Housekeeping genes as internal standards: use and limits. *J Biotechnol* 75: 291-295,

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

WHO (2012), Guidance for immunotoxicity risk assessment for chemicals. WHO press, Paris,

Winer, J., Jung, C.K., Shackel, I., et al. (1999), Development and validation of real-time

quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. *Anal Biochem* 270: 41-49, 10.1006/abio.1999.4085

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 sulfoxide
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
I.I.-SLR-LA : Inhibition index of SLR-LA
IL-2 : Interleukin-2
IL-8 : Interleukin-8

JaCVAM : the Japanese Center for the Validation of Alternative Methods
JSIT : Japanese Society of Immunotoxicology
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-ImmunoTox Assay
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
nSLG-LA : normalized SLG luciferase activity
nSLO-LA : normalized SLO luciferase activity
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
SLG-LA : SLG luciferase activity
SLO : Stable luciferase orange

SLO-LA : SLO luciferase activity

SLR : Stable luciferase red

SLR-LA : 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 : the United Nations Globally Harmonized System of Classification and Labeling of Chemicals

VMT : Validation Management Team