4th Meeting for the MITA Validation Study

October 4–5th, 2018
Kobe University, Kobe, Japan

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

	October 4
	Welcome address and housekeeping
Ashikaga:	We have postponed the discussion about reproducibility until tomorrow.
Aiba:	Thank you for being here today. This is an important meeting for our validation study. We have found that the problems we face are really very difficult yet also interesting, and I hope will be able to reach agreement at this meeting.
Kojima:	Today's discussion will focus on predictive capacity. And Dr. Aiba has a new proposal. Tomorrow we will discuss reproducibility, the study plan, and other issues.
	Report of draft validation report and discussion of predictive capacity
Aiba:	(PowerPoint presentation)
Corsini:	We need to look at all data that shows T-cell response to chemicals for which we have mechanistic information and be aware that some of those mechanisms might be affecting other things. And compare this with any other data for which we know the response to these chemicals.
Germolec:	Data from the IL2 assay seems to suggest that there are chemicals that are immunosuppressive but probably are not targeting T-cells. This also shows the power of the combined assays to enable predictions where individual assays give only equivocal results. Perhaps we can rethink the definition of immunotoxic and specify that this assay is predictive of chemicals that target T-cells, and it becomes a tool in the toolbox.
Inoue:	Cytokine release and cell proliferation are important for judging T-cell dependent mechanisms.
Aiba:	That will be added.
Corsini:	What are the criteria for the tentative designation of immunotoxic chemicals?
Aiba:	We look at different effects such as changes in the thymus weight, proliferation, serum immunoglobin.
Germolec:	I think that the observed endpoints are more important than the number of assays that give positive or negative results. So, for example, if we have relevant human epidemiology data, that is all would need to classify based on human health effects. It is good to have mechanistic and other data, but if you are doing risk assessment, all you need is evidence of an effect on a human population.
Inoue:	Very few chemicals show immuno-augmentation in in vivo studies. So even if IL2 shows augmentation, that might not be corroborated by in vivo studies.
Aiba:	We are going to classify 25 chemicals for potency and target?
Germolec:	For this assay, the first thing we should look at is whether it targets T-cells or not. That is its strength and where the focus should be. Then we can discuss other immune effects or potency and other things.
Aiba:	Chemical Cytokine production (in vivo, ex vivo) Mitogen response DTH TDAR T-cell immunotoxic Non T-cell immunotoxic None
Germolec:	If we have trouble finding chemicals with in vivo T-cell data, we might consider running a small set of chemicals on animals.
Kojima:	How many chemicals have you tested in-house?

Aiba:	60. Here is a list of chemicals we have published, and many of them are drugs, so the effects are well known. The criteria used here is slightly different, however.
Corsini:	So, you could apply the new criteria to this data.
Germolec:	Instead of saying "T-cell immunotoxic or non-immunotoxic," perhaps we should say "Targets T-cells: Yes or No."
Corsini:	But it is difficult to determine a direction, because there will be in vivo data that sometimes show changes in both directions.
Germolec:	So, we should focus just on whether it affects T-cells or not.
	October 5
	Discussion of Reproducibility
Omori:	We used criteria 5 shown on page 30 of section 7-2. Recalculation was done for Phase I and Phase II. For Phase I, WLR was 80%, 100%, and 80% for the three laboratories, and BLR was 80% for five chemicals. For Phase II, BLR was 80% for twenty chemicals.
Aiba:	We found a problem in the calculation sheet but applied a correction to get these figures.
Corsini:	A concordance of 80% is a reasonable level. Even if there was a mistake in the calculation sheet, once it was corrected, the figures are what we expected, so I think we can all agree that the method achieves reproducibility targets.
Kojima:	
	Conclusions
Corsini:	Yesterday we discussed how we cannot expect accurate predictions for chemicals that do not target T-cells, so we have 25 chemicals that we reviewed to determine whether or not they target T-cells.
Aiba:	We established criteria and reviewed the 25 chemicals yesterday, but there are some points we need to discuss.
Corsini:	If we have a chemical that we consider undetermined but there are in vivo or other test results that suggest it does target T-cells, should we include it?
Germolec:	There are some chemicals for which other tests find both suppression and augmentation, so I think that when we see the same thing in IL2 results, it is an indication of how sensitive this test is.
Roggen:	I wonder if we should go back to look at the ones that are undetermined because there was no relevant data.
Germolec:	I think we need to distinguish between cases where there is not enough data to determine whether or not it targets T-cells and cases where there is evidence that other types of cells are also targeted.
Corsini:	So, results for chemicals that are undetermined should not be included in the discussion of predictive capacity.
Aiba:	So, at the moment, our predictive capacity is 16/24 or 67%.
Corsini:	How does that compare with your earlier in-house data? Although you probably should apply the current criteria to make a relevant comparison.
Aiba:	I think that drugs that we know target T-cells can be predicted accurately, but I am not so sure that other chemicals with effects that are not so well known can be predicted.
Corsini:	We won't be able to make any conclusions using the in-house data until we do a further review of the literature and then apply criteria 5. And it would be helpful in determining an applicability domain to be able to provide a mechanistic explanation for chemicals that give IL2 results that aren't concordant with results from other test methods.
Corsini:	It is important that accuracy be greater than 50%, but an accuracy of 65 or 70% is fine for one tool in the toolbox
Roggen:	I wonder if it might not make the test more useful just to forget about suppression and augmentation and focus on saying yes or no for targeting the biological function of T-cells. Aspects that are concentration-dependent are not so useful, because concentrations used in vitro are not necessarily relevant in vivo.
Germolec:	Many of the in vitro tests currently being used are only looking for biological effects. They aren't concerned with whether the effect is up or down, but rather is there an effect at all.
	Plan for a peer review
Kojima:	Here is my plan for the peer review. (PowerPoint presentation)
Aiba:	I hope to finish the first draft validation report by the end of November.
Kojima:	We would like to fix the PRP members by the end of November, and early discussions will be by email and a teleconference if necessary. And a F2F meeting at the end of February 2019. And the have the report by the end of July.
Corsini:	I suggest Henk Van Loveren or Marc Pallardy for the EU expert.

Germolec:	Mike Luster, Madeline Fort, or Haley LaNef Ford for the US expert. And for the Korean expert, Sanghyun Kim from Korea.
Kojima:	I would like to propose that Japan submit a SPSF, and if it is approved, then I will coordinate a DRP with the VMT and the PRP and would like to discuss the TOC next July. The target is the WNT meeting in April 2020.
	Validation plan for the IL-1β assay
	Introduction and Protocol
Aiba:	Please look at page 5 of the file Multi-Immuno Tox Assay protocol for-TGCHAC-A4 ver. 007E 20180712. This shows the protocol for the assay. And on page 36, the criteria are described. Acceptance criteria will be determined based on Phase 0 results.
Inoue:	What about the passages of the cells before the assay?
Aiba:	Cell cultures requirements are described on page 10.
Corsini:	Do the cell lines and reagents need to be purchased from a specific supplier? Do you provide a catalog number? In the future, will these materials be available from several sources?
Aiba:	In the future, yes, although we tried several different LPS and settled on this one.
Roggen:	It would be good to describe the different LPSs in terms of stimulation potency and show what kind of differences can be expected.
Corsini:	We should keep in mind the potential to replace bovine calf serum with human serum.
Aiba:	The prediction criteria are two or more consecutive statistically significant positive (negative) data or one statistically significant positive (negative) data with a trend in which at least 3 consecutive data increase (decrease) in a dose dependent manner.
Corsini:	So, this is the same as Criteria 5.
Roggen:	Again, it should be focused on yes or no.
Germolec:	The criteria need to define suppression and augmentation, but then the data analysis can be based on yes or no.
Aiba:	We used different plates which show different backgrounds of DMSO. Also, we thought about what kind of counts should be expected. Three of four laboratories achieved similar stimulation with LPS. So, we think that a 5-fold or greater induction of FinSLG-LA should be included in the acceptance criteria.
Omori:	Do you have any ideas how to increase the fold induction at the lab that was low?
Aiba:	The lab that was low needs to increase its fold induction the same level as the other labs before starting Phase I.
Corsini:	Did all the participating laboratories have materials from the same manufacturing lots?
Aiba:	Yes.
Corsini:	Were the readers all calibrated to the same standards, as well?
Aiba:	Yes.
Germolec:	Are there any other markers of proliferation that could be used to understand whether or not they are all getting the same levels of activation? What will you do in the next phase so to improve this situation for that lab?
Aiba:	We will investigate the situation to find out why they had lower fold induction.
Roggen:	When you saw there was a problem, did you send them new cells to use?
Aiba:	No, we just did the one test. But we do understand that this is an issue that we have to correct before we proceed.
	Results of the Pre-Validation Study
Omori:	Criteria 5 is from IL2 and these graphs have reference lines at ±35%. There are a number of tests, however, in which changing the cutoff value from 35% to 20% would change the results to suppression. The last sheet summarizes the results using 35%, 20%, and 25% cutoff values.
Inoue:	Please explain the why the judgement is suppression rather than augmentation on page 9.
Omori:	If the confidence interval had extended below the zero line then it could be considered a trend, but it did not.
Germolec:	Even though the response was not as high as some of the data from the other labs, the lab that did not perform as well showed a similar trend as the results from the other labs.
Roggen:	There are two things here. The ratio is important to help people using a different machine to define a window, but then an absolute value can be used to make the judgment.
Germolec:	In terms of stimulation, you need a cutoff for DMSO. And it looks like 5 is too low but 10 is two high. So, we

	need to determine what fold induction is sufficient to ensure optimal LPS response.
Corsini:	The readers are the same and the materials are the same, so there needs to be an explanation of why one lab had significantly lower figures.
Aiba:	Unfortunately, we changed the plates before the Phase 0 study. So, it is difficult to determine what was different based on in-house data.
Corsini:	Perhaps viability is one explanation for why there was variation in the response.
Inoue:	Perhaps the time since the passage affected the response.
Aiba:	This is stipulated in the protocol.
Roggen:	One shows high response with large variation, but another shows low response but small variation. It's possible that different techniques cause some of these variations. There are many variables, even the age of the equipment could be a factor.
Aiba:	We agree that we have to devise criteria to ensure a sufficient fold induction and find a reason and a countermeasure for this problem. Maybe we need to redo some tests.
Yamakage:	Well, it will take some time to do more tests, but first we need to resolve our problem.
Germolec:	Stimulation by LPS should be independent of the plate being used.
Aiba:	Yes, we will examine our data.
Roggen:	If there is no historical data, then you will need to do some additional testing to acquire data you can use.
Germolec:	Looking at the dates, it appears that the response was dropping in the later experiments, and since FDSC performed their tests last, perhaps their low counts were related to that.
	Study Plan for the Validation Study
Kojima:	This is the same study plan as for IL2 and calls for 25 test chemicals: 5 in Phase I and 20 in Phase II. Is this workable for the participating laboratories?
Yamakage:	It might be difficult.
Kojima:	We need to start the Phase I testing by January. The timeline calls for starting in November, but we have to find a solution to the Phase 0 problem, so I think the start date will be January. Do we need a teleconference in December before starting Phase I?
Roggen:	We need to see a list of what has been done to solve the problem.
Germolec:	And I think email might be better, because December is a busy time of year to try to arrange a teleconference.
Aiba:	I think the problem is clear and we should be able to find a solution.
	Comments from the Participating Laboratories
Yamakage:	We don't have any clear ideas for solving the problem, so I will need some time to check.
Roggen:	It takes time to identify problems in reproducibility. Sometimes it is something as simple as differences in pipette technique, so there are many things to consider and it is difficult to plan beforehand. You have to stay openminded and compare one thing after the other.
Germolec:	But it also needs to be noted that if the protocol is that sensitive to the techniques used, then its transferability to other laboratories will be in question.
Corsini:	In a situation where only one out of four labs is getting different results, then you have to ask why.