

## 厚生労働科学研究補助金（化学物質リスク事業）

### 化学物質の動物個体レベルの免疫毒性データ集積とそれに基づく Multi-ImmunoTox assay (MITA) による予測性試験法の確立と国際標準化

分担研究報告書（平成 30 年、令和元年、令和 2 年度総合）

分担研究者：大森崇

#### 研究要旨

**【背景と目的】** Multi-Immuno Tox Assay(以下、MITA)は、化学物質がヒトの免疫系に与える影響を *in vitro* で評価することを目標とする試験法であり、化学物質の IL2、IL1 $\beta$  の発現をとらえる IL-2 Luc アッセイ及び IL-1 $\beta$  Luc アッセイが開発された。これら 2 つのアッセイの判定アルゴリズムの開発を行うとともに、*in vivo* に対する予測性の評価を目的とした。

**【方法】** バリデーション研究を通して提案された判定方法について、アルゴリズムを作成した。また、63 物質を用いて、*in vivo* の判定結果と IL-2 Luc アッセイ及び IL-1 $\beta$  Luc アッセイの判定結果を用いた 5 つの方法について、Receiver Operation Characteristics 曲線の曲線下面積である Area Under the Curve (AUC) の値とその 95%信頼区間とともに感度、特異度を計算した。

**【結果】** アルゴリズムの開発では、2 つ前からの濃度における反応の指標とその 95%信頼区間、カットオフ値を用いて複数のステップからなる判定アルゴリズムを開発した。信頼性の評価では、5 つの方法のうち、最も AUC が高かった方法は 2 つのアッセイを組み合わせた方法において、AUC は 0.59 (95%信頼区間 [0.44, 0.75])、感度 77.6%、特異度 14.3%であった。

**【結論】** アルゴリズムにより定式化することで、IL-2 Luc アッセイ及び IL-1 $\beta$  Luc アッセイについて、濃度反応関係の図を参照しながら基準に照らし合わせていた判定について、広く実装を可能とした。

信頼性の検討では、検討した物質において高い信頼性の結果は得られなかったが、結果は暫定的であり、評価を行う物資を含めて、信頼性の評価のためにさらなる検討が必要であることが課題となることがわかった。

#### A. 研究目的

Multi-Immuno Tox Assay(以下、MITA)は、化学物質がヒトの免疫系に与える影響を *in vitro* で評価することを目標とする試験法である。MITA では、ある化学物質の IL2、IL1 $\beta$  の発現をとらえる IL-2 Luc アッセイ及び IL-1 $\beta$  Luc アッセイが開発され、経済協力開発機構 (Organisation for Economic Co-operation and Development : OECD) の試験

法ガイドライン (Test Guideline : TG) としての公定化を目指し、バリデーション研究が施行されている。

IL-2 Luc アッセイおよび IL-1 $\beta$  Luc アッセイでは、特定の化学物質の発現を評価は独立した 3 回の実験から得られる測定値を用いて行われる。バリデーション研究を行う中で、このアッセイに関する実験方法だけでなく、判定方法についても検

討が行われてきた。判定方法については、

- ・ 特定のある一濃度で発現が生じることがあるため、濃度別の統計的有意性のみで判定を行うと高感度になりすぎる
- ・ 濃度依存性をある程度考慮したい
- ・ 他の試験法の判定で行われてきたように 3 回の実験の判定は独立に行うことが望ましい
- ・ 特定の濃度以上は判定を行わない

という意見を出し合い判定方法が決定された。

2 つのアッセイは、測定値から各濃度別に計算される %suppression という指標についてその点推定値と 95%信頼区間とともに、カットオフ値と呼ばれる参照線を引いた図から読み取り判定を行う。このような判定は、濃度に対するその反応の図を必ず確認することになるという利点があるが、基準が複雑で誤りを犯しやすいという欠点も有する。初年度は、IL-2 Luc アッセイ、2 年度目は IL-1 β Luc アッセイの判定アルゴリズムの開発を行った。

新たに開発された試験法が特定の毒性の判定を行うことが可能かどうかの判断を行う際に、再現性と予測性という異なる 2 種類の評価が必要になる。再現性は、同一の物質を同一の試験方法で適用したときの施設内再現性、施設間再現性の評価が求められる。予測性は、ターゲットとする毒性の判定と新たに開発された試験法の判定がどの程度一致するかの評価が求められる。最終年度は、IL-2 Luc アッセイ及び IL-1 β Luc アッセイ単独および組み合わせた際の予測性についての評価を目的とした。

## B. 研究方法

### B.1. 判定アルゴリズムの開発

#### IL-2 Luc アッセイ及び IL-1 β Luc アッセイのデータの特徴

IL-2 Luc アッセイ及び IL-1 β Luc アッセイでは、独立に実施された 3 回の実験が行われる。個々

の実験は濃度 0(DMSO)群と各濃度群との対比較により、判定を行う。1 回の実験には、96 穴プレートが用いられ、溶媒を用いた濃度 0 と濃度 1~濃度 10 までの計 11 段階の濃度群の測定値が得られる。個々の濃度群では 4 回の繰り返しがある。これらの測定値から、濃度ごとに、濃度 0 に対する被験物質の抑制割合を示す %suppression という 1 つの要約指標を計算できる。測定値に繰り返しがあるため、誤差的な変動を考慮して 95%信頼区間を構成できる。

#### 各実験から得られる測定値と指標について

1 回の実験において、96 穴プレートの各セルから SLG-LA (SLG ルシフェラーゼ活性)、SLO-LA (SLO ルシフェラーゼ活性)、SLR-LA (SLR ルシフェラーゼ活性) の 3 種類の発光に関する測定値が得られる。化学物質の評価において、第 i 番目の濃度 (i=0,1,2,...,10) の第 j 番目の繰り返しの測定値をそれぞれ SLG-LA<sub>ij</sub>、SLO-LA<sub>ij</sub>、SLR-LA<sub>ij</sub> とする。

これらの測定値を用いて、判定のための指標が得られる。

$$\begin{aligned} \text{I.I. SLR-LA}_i &= (\overline{\text{SLR-LA}}_i) / (\overline{\text{SLR-LA}}_0), \\ \text{nSLG-LA}_{ij} &= \text{SLG-LA}_{ij} / \text{SLR-LA}_{ij}, \\ \% \text{suppression}_i &= \left\{ 1 - \left( \overline{\text{nSLG-LA}}_i \right) / \left( \overline{\text{nSLG-LA}}_0 \right) \right\} \times 100, \end{aligned}$$

ただし、

$$\begin{aligned} \overline{\text{SLG-LA}}_i &= \sum_{j=1}^n \text{SLG-LA}_{ij} / n_i, \\ \overline{\text{SLR-LA}}_i &= \sum_{j=1}^n \text{SLR-LA}_{ij} / n_i \end{aligned}$$

であり、 $n_i$  は第 i 濃度の繰り返し数であり実質 4 である。

%suppression<sub>i</sub> は、第 i 濃度での免疫毒性の程度を %換算した値である。この指標は基本的には平

均値の比であるため、デルタ法を用いてその 95%信頼区間を構成することができる。

は細胞の状態が正常かどうかを判断するための指標である。

### IL-2 Luc アッセイの実験ごとの判定

IL-2 Luc アッセイの最終的な判定は各実験の判定である「Suppression」「Augmentation」「No effect」に基づき行われる。各実験の免疫毒性ありの判定は以下の 3 つの基準をすべて満たした場合からなる。

1. ある濃度の%suppression の値がカットオフ値上限以上でかつ 95%信頼区間の下限が 0 よりも大きい、もしくは ある濃度の%suppression がカットオフ値下限以下で 95%信頼区間の上限が 0 よりも小さい。(1)

2. (1)に対して

隣り合ういずれかの 2 濃度で 95%信頼区間の下限が 0 より大きくそれぞれの%suppression の値がカットオフの上限より大きくなるか、95%信頼区間の上限が 0 より小さくでそれぞれの%suppression の値がカットオフの上限より小さくなる。(2-1) または、ある 1 濃度で 95%信頼区間の下限が 0 より大きくなる場合にその濃度を含めた続く 3 濃度の%suppression の値が増加傾向を示すか、95%信頼区間の上限が 0 より小さくなる場合にその濃度を含めた続く 3 濃度の%suppression の値が減少傾向を示す。ただし、この場合、%suppression の値が 0 をまたいでよいのは 1 濃度のみであり、0 をまたいだ濃度の 95%信頼区間の上限が 0 以下にならないもしくは下限が 0 以上にならない。(2-2)

3. I.I.-SLR-LA が 0.05 以上となる濃度が判定では有効となる濃度である。

### IL-1β Luc アッセイの実験ごとの判定

アッセイの最終的な判定は各実験の判定である

「Suppression」か「No effect」に基づき行われる。各実験の免疫毒性ありの判定は以下の 3 つの基準をすべて満たした場合からなる。

1. ある濃度の%suppression の値がカットオフ値上限以上でかつ 95%信頼区間の下限が 0 よりも大きい、もしくは ある濃度の%suppression がカットオフ値下限以下で 95%信頼区間の上限が 0 よりも小さい。(1)

2. (1)に対して

隣り合ういずれかの 2 濃度で 95%信頼区間の下限が 0 より大きくそれぞれの%suppression の値がカットオフの上限より大きくなるか、95%信頼区間の上限が 0 より小さくでそれぞれの%suppression の値がカットオフの上限より小さくなる。(2-1) または、ある 1 濃度で 95%信頼区間の下限が 0 より大きくなる場合にその濃度を含めた続く 3 濃度の%suppression の値が増加傾向を示すか、95%信頼区間の上限が 0 より小さくなる場合にその濃度を含めた続く 3 濃度の%suppression の値が減少傾向を示す。ただし、この場合、%suppression の値が 0 をまたいでよいのは 1 濃度のみであり、0 をまたいだ濃度の 95%信頼区間の上限が 0 以下にならないもしくは下限が 0 以上にならない。(2-2)

3. I.I.-SLR-LA が 0.05 以上となる濃度が判定では有効となる濃度である。

## **B.2. 予測性の検討**

### 利用したデータ

検討に利用したデータは、1 施設で実施された IL-2 Luc アッセイ及び IL-1β Luc アッセイの結果及び、文献により集められた *in vivo* の判定結果を含めた 63 物質（陽性物質 49 物質、陰性物質 14 物質）のデータある。

### 検討した判定方法

予測性の評価は、

1. IL-2 Luc アッセイ単独の判定
2. IL-1β Luc アッセイ単独の判定
3. IL-2 Luc アッセイと IL-1β Luc アッセイ単独がともに陽性と判定した場合に陽性、そうでない場合に陰性とした判定
4. IL-2 Luc アッセイと IL-1β Luc アッセイ単独のどちらか陽性と判定した場合に陽性、そうでない場合に陰性とした判定
5. ロジスティック回帰分析の説明変数に IL-2 Luc アッセイの判定結果と IL-1β Luc アッセイ単独の判定結果を含めて予測した場合の判定とした。

### 効果の評価指標

上記の 1~5 について、効果の評価指標は、Receiver Operation Characteristics (ROC) 曲線の曲線下面積である Area Under the Curve (AUC) の値とその 95%信頼区間とともに感度、特異度とした。上記の方法 5 についての感度と特異度を求めるカットオフ値は、感度が 100%未満の最大値となる値とした。

### データの取り扱い

上記の評価において、IL-2 Luc アッセイの判定は、immunosurpression、immunoaugmentation を「陽性」として、no effect を「陰性」とした。また、IL-1β Luc アッセイについては、immunosurpression、を「陽性」として、no effect を「陰性」とした。

## C. 研究結果

### C.1. 判定アルゴリズムの開発

#### 判定アルゴリズムの前提

基準(3)に対して有効な濃度の測定値を用いることにする。

基準 (2-2) では、続く 3 濃度の%suppression の

値の大小関係が必要となるため、同時に 3 濃度を比べる必要がある。第 i 濃度の%suppression である%suppression<sub>i</sub>に対して、1 つ前の濃度の値を%suppression<sub>i</sub><sup>(-1)</sup>、2 つ前の濃度の値を%suppression<sub>i</sub><sup>(-2)</sup>とする。

基準 (2.1) では、隣り合う 2 つの%suppression の 95%信頼区間の比較を行うことになる。%suppression<sub>i</sub>の 95%信頼区間の下限を、上限をとし、これらの 95%信頼区間に対して、一つ前の濃度の下限を、上限をUpper<sub>i</sub>、2 つ前の濃度の下限をLower<sub>i</sub><sup>(-1)</sup>、上限をUpper<sub>i</sub><sup>(-2)</sup>とする。

また、カットオフ値の上限 (Suppression に関係する) を、カットオフ値の下限 (Augmentation に関係する) をとする。

### IL-2 Luc アッセイの判定アルゴリズム

#### Step 1

全ての濃度の となる を欠測とする。

#### Step 2-1

以下の条件を満たした場合に新しい変数 TwoSig-S<sub>i</sub>を 1、そうでない場合は 0 とする。

$$\text{Lower}_i^{(-1)} > 0 \text{ かつ } \text{Lower}_i > 0$$

#### Step 2-2

以下の条件を満たした場合に新しい変数 TwoSig-A<sub>i</sub>を 1、そうでない場合は 0 とする。

$$\text{Upper}_i^{(-1)} < 0 \text{ かつ } \text{Upper}_i < 0$$

#### Step 3-1

以下の条件を満たした場合に新しい変数 Trend-S<sub>i</sub>を 1、そうでない場合は 0 とする。

$$\%suppression_i^{(-1)} > 0 \text{ かつ } \%suppression_i > 0$$

かつ

$$\%suppression_i^{(-2)} < \%suppression_i^{(-1)}$$

かつ

$$\%suppression_i^{(-1)} < \%suppression_i$$

かつ

$$\text{Upper}_i^{(-2)} > 0$$

### Step 3-2

以下の条件を満たした場合に新しい変数 **Trend-A<sub>i</sub>** を 1、そうでない場合は 0 とする。

$$\%suppression_i^{(-1)} < 0 \text{ かつ } \%suppression_i < 0$$

かつ

$$\%suppression_i^{(-2)} > \%suppression_i^{(-1)}$$

かつ

$$\%suppression_i^{(-1)} > \%suppression_i$$

かつ

$$Lower_i^{(-2)} < 0$$

### Step 4-1

以下の条件を満たした場合に新しい変数 **IndConc-S<sub>i</sub>** を 1、そうでない場合は 0 とする。

$$\{TwoSig-S_i = 1 \text{ かつ } \%suppression_i > Cut^{(S)}\}$$

$$\text{かつ } \%suppression_i^{(-1)} > Cut^{(S)}\}$$

または

$$\{Trend-S_i = 1 \text{ かつ } Lower_i > 0$$

$$\text{かつ } \%suppression_i > Cut^{(S)}\}$$

### Step 4-2

以下の条件を満たした場合に新しい変数 **IndConc-A<sub>i</sub>** を 1、そうでない場合は 0 とする。

$$\{TwoSig-A_i = 1 \text{ かつ } \%suppression_i < Cut^{(A)}\}$$

$$\text{かつ } \%suppression_i^{(-1)} < Cut^{(A)}\}$$

または

$$\{Trend-A_i = 1 \text{ かつ } Upper_i < 0$$

$$\text{かつ } \%suppression_i < Cut^{(A)}\}$$

### Step 5

いずれかの濃度で **IndConc-S<sub>i</sub> = 1** の場合に

Suppression と判断する。

いずれかの濃度で **IndConc-A<sub>i</sub> = 1** の場合に

Augmentation と判断する。

それ以外の場合には No effect とする。

## II-1 β Luc アッセイの判定アルゴリズム

### Step 1

濃度が 2000 を超えるデータを除外する。

全ての濃度の となる を欠測とする。

### Step 2

以下の条件を満たした場合に新しい変数 **TwoSig-S<sub>i</sub>** を 1、そうでない場合は 0 とする。

$$Lower_i^{(-1)} > 0 \text{ かつ } Lower_i > 0$$

### Step 3

以下の条件を満たした場合に新しい変数 を 1、そうでない場合は 0 とする。

$$\%suppression_i^{(-1)} > 0 \text{ かつ } \%suppression_i > 0$$

かつ

$$\%suppression_i^{(-2)} < \%suppression_i^{(-1)}$$

かつ

$$\%suppression_i^{(-2)} < \%suppression_i^{(-1)}$$

かつ

$$\%suppression_i^{(-1)} < \%suppression_i$$

### Step 4

以下の条件を満たした場合に新しい変数 **IndConc-S<sub>i</sub>** を 1、そうでない場合は 0 とする。

$$\{TwoSig-S_i = 1 \text{ かつ } \%suppression_i > Cut^{(S)}\}$$

$$\text{かつ } \%suppression_i^{(-1)} > Cut^{(S)}\}$$

または

$$\{Trend-S_i = 1 \text{ かつ } Lower_i > 0$$

$$\text{かつ } \%suppression_i > Cut^{(S)}\}$$

### Step 5

いずれかの濃度で **IndConc-S<sub>i</sub> = 1** の場合に Suppression (positive) と判断する。

それ以外の場合には No effect とする。

## **C.2. 予測性の検討**

5つの方法のそれぞれにおける ROC 曲線の曲線下面積である AUC とその 95%信頼区間を表 1 に示す。いずれの方法においても AUC が 0.6 より大きな値となるものはなく、95%信頼区間の下限は 0.5 を下回っていた。

表 1 5つの方法の AUC

方法	AUC	95% Confidence	
		Lower	Upper
1	0.56	0.45	0.68
2	0.57	0.42	0.71
3	0.59	0.44	0.73
4	0.54	0.43	0.65
5	0.59	0.44	0.75

表 2 に 5 つの方法の感度および特異度を示す。80%以上の感度を得ることができなかった。

表 2 5 つの方法の感度 (%) と特異度 (%)

方法 5 は感度が 1 未満の最大となるようにカットオフ値を決めた場合

方法	感度 (%)	特異度 (%)
1	73.5	14.3
2	51.0	35.7
3	46.9	35.7
4	77.6	14.3
5	77.6	14.3

## D. 考察

### D.1. 判定アルゴリズムの開発

MITA の IL-2 Luc アッセイ、IL-1 $\beta$  Luc アッセイの施設内および施設間再現のバリデーションは終了しており、良好な再現性が得られている。これまで、アッセイの免疫毒性の判定は濃度反応曲線のグラフから読み取ることで行われている。グラフから瞬時に Suppression、Augmentation、No effect であるかを判断することは難しい。現在は、図を描くことで複数の目でチェックを行いながらバリデーション研究を行っているが、判定結果を誤る可能性が否定できない。

これらのアッセイが広く利用するようになる際には、本研究で検討したアルゴリズムをデータシートに組み込むことで、これまで費やされてきた

多くの負担を減らすことができるであろう。

### D.2. 予測性の検討

予測性の検討では、IL-2 Luc アッセイ及び IL-1 $\beta$  Luc アッセイの予測性の検討を行った。予測性は *in vivo* の結果と *in vitro* の関連の強さを評価する指標である。関連が見いだせず、偶然誤差の変動である場合、AUC は 0.5 となることが知られている。つまり、AUC が 0.5 に近い評価結果で高い予測性を有するとはいえないことになる。AUC の 95%信頼区間が 0.5 を含むことは、本研究で検討したデータセットにおいて、AUC が 0.5 であることを否定できないことを意味する。評価した 5 つの方法のいずれにおいても 95%信頼区間が 0.5 を含んでいた。AUC は全体の評価を与えるが、感度と特異度は、*in vivo* の結果が陽性の場合と陰性の場合のそれぞれについての予測能を示す指標である。どの方法であっても、感度は 80%を超えることはなく、特異度は 50%に満たなかった。検討した 5 つの方法のうち、方法 1 が IL-2 Luc アッセイ単独であり、方法 3 から 5 が IL-2 Luc アッセイに IL-1 $\beta$  Luc アッセイを組み合わせた方法となる。したがって、本検討の範囲に関しては、IL-2 Luc アッセイと IL-1 $\beta$  Luc アッセイを組み合わせても、予測性の改善はあまり大きくないといえるであろう。

しかしながら、ここでの検討結果のみから 2 つアッセイの予測性が悪いと結論づけるべきではないと思われる。特異度が特に低かった理由は、2 つアッセイの結果が陽性と判定された物質が多かったことによる。陽性物質が 49 であったのに対し、陰性物質はわずか 14 物質である。このうちの多くの物質が 2 つのアッセイで陽性と判定されていた。予測性の性能評価の前に行われた再現性の検討では、2 つのアッセイとも、陰性物質は良好な施設間再現性の下、用量が増加しても反応を示さなかつ

た。したがって、2つのアッセイがどのような物質にも反応を示すアッセイであるとは考えにくい。より広い化学物質に対して、これらのアッセイを適用した場合の結果の特異度は異なるものになる可能性がある。*in vivo*の陽性となる物質が何を検出しているのかということもさらなる検討が必要であると思われるが、検討に用いた63物質より広い化学物質の中で*in vivo*で確認するまでもないとして評価されてこなかった物質は、検討した物質には含めることができているだけなのかもしれない。

感度に注目した場合、感度も決して高いという結果ではなかった。これらは過去2年で検討した判定方法のカットオフ値と関係があるかもしれない。

本検討の範囲には含まれていないが、IL-2 Luc アッセイ及びIL-1 $\beta$  Luc アッセイのどちらも高い施設内、施設間再現性を示しているアッセイである。本検討から、予測能は現段階ではまだ試験法の課題を残しているといえるが、より広い物質を含めたさらなる検討が必要である。

#### E. 結論

IL-2 Luc アッセイ及びIL-1 $\beta$  Luc アッセイの判定アルゴリズムを開発した。これらのアッセイの信頼性の検討を行った結果、信頼性の評価のためにさらなる検討が必要であることがわかった。

#### F. 健康危険情報

なし。

#### G. 研究発表

Kimura Y, Yasuno R, Watanabe M, Kobayashi M, Iwaki T, Fujimura C, Ohmiya Y, Yamakage K, Nakajima Y, Kobayashi M, Mashimo N, Takagi Y, Omori T, Corsini E, Germolec D, Inoue T, Rogen EL, Kojima H, Aiba S. 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. *Toxicol In Vitro*. 2020. doi: 10.1016/j.tiv.2020.104832.

#### H. 知的財産権の出願・登録状況

なし。

#### 参考文献

- 1) Delwiche, L. D. and Slaughter, J. S. The little SAS book a primer 5th ed. SAS, 2012.
- 2) Paul, D. A. Logistic Regression Using SAS: Theory and Application, 2nd ed, SAS, 2012.