（別紙）研究成果の概要（英文）

Title of research project	Study on the development of new evaluation support technology: Examination of database utilization method for toxicity prediction
Research project number	（1801）
Research period	FY 2018 – 2019
Name of principal research investigator (PI)	Masahiro Tohkin

Abstract/Summary

It is desired to establish the *in silico* prediction methods that predict the toxicity of chemical substances using open database of the toxicity test information. However, the prediction of the repeated-dose toxicity (RDT) has difficulties because of the complexity of mechanism, the diversity of chemical structure, and the lack of information on toxicity test. Therefore, the practical application of *in silico* prediction has not been achieved. The aims of this study were to examine the optimal toxicity findings and necessary data for *in silico* prediction in the RDT studies, and to clarify the necessary information to be added to the toxicity database and the characteristics of the prediction tools. First, we extracted toxicological findings from the toxicity database for the *in silico* prediction model construction. Next, we conducted a case study of the (Q) SAR prediction model and the category approach prediction model using the extracted toxicological findings from the toxicity database. Finally, we verified the toxicity database and evaluation support tools suitable for *in silico* evaluation.

We extracted toxicological findings from rat hepatotoxicity, nephrotoxicity, and hematologic toxicity using the RDT study database of HESS. For each toxicological finding, we created a data set for *in silico* evaluation, which included the presence or absence of toxicity and the structural information of chemical substances. This dataset has been organized using FileMaker and will be useful as a database basis for *in silico* evaluation.

Next, we conducted a case study of the prediction of hepatotoxicity, nephrotoxicity, and hematologic toxicity from molecular descriptors using (Q) SAR model based on machine learning. For the hepatotoxicity, we constructed the prediction model by random forest (RF) for "increased liver weight and histopathological findings", and obtained the high prediction performance with MCC 0.56 and AUC 0.84. In the case of nephrotoxicity, the model of "increased kidney weight and histopathological findings" showed high predictive performance of MCC 0.63 and AUC 0.89. It was suggested that a high-performance model could be constructed by properly combining toxicological findings. In addition, the prediction performance was significantly improved in the models of "ALT rise or AST rise" and "Cre rise or BUN rise" due to the elimination of data imbalance by SMOTE. Furthermore, by

setting the applicability domain, we have constructed a model with high generalization ability that can maintain prediction performance even with external data.

Using the HESS RDT study database, we evaluated a novel statistical read-across method, in which the similarity among chemicals were calculated using various sets of molecular descriptors. In this evaluation, we selected major RDT endpoints related to liver and kidney toxicity and anemia. The number of the compounds evaluated were limited because of the introduction of thresholds for the similarity judgement. The results indicate that the dataset and descriptors used might be applicable to the evaluation of anemia-related endpoints. This is partly because the dataset contains a number of compounds related to known anemia-inducing anilines and phenols. The results also indicate that even for a given endpoint the accuracy of toxicity evaluation depended on the set of molecular descriptors and that even when the same descriptor set was used, the evaluation accuracy varied depending on the target endpoint. Taken together, we have demonstrated that it is important to select appropriate descriptors and to set an appropriate threshold for similarity judgement depending on datasets and target endpoints in our statistical read-across approaches.

Finally, we verified the toxicity database and evaluation support tools suitable for *in silico* evaluation from the results of the case study. Regarding the toxicity database for *in silico* evaluation, it is important to obtain the data on substances that have toxic effects (positive substances). In addition, it is necessary to prepare a database that contains as many substances as possible, and structure-limited database that contains many substance groups with specific partial structures in order to carry out highly objective read-across. With regard to the evaluation support tools, the prediction model by (Q) SAR using RF as machine learning enabled high-performance prediction by devising to properly process the data. In addition, it was shown that the HESS database and specific molecular descriptors may be useful for the read-across of hematologic toxicity associated with anemia in the novel *in silico* read-across method using statistical methods.

The results of this research provides the criteria for the indispensable toxicity dataset and the evaluation support tools in the food safety risk assessment, and these information are very important information for *in silico* evaluation method.

1. List of papers published on the basis of this research

1) Ambe K, Sakakibara Y, Sakabe A, Makino H, Ochibe T, Tohkin M. Comparison of the developmental/reproductive toxicity and hepatotoxicity of phthalate esters in rats using an open toxicity data source. ◎ *J. Toxicol. Sci.* **44**, 245-255 (2019).

２．List of presentations based on this research

1) Tatsuya Ochibe、Kaori Ambe、Masahiro Tohkin: Machine learning models for predicting hepatotoxicity and renal toxicity based on HESS toxicological test database. NCU Contact Points in Asia Symposium 2019 (2019.12) (Nagoya)
2) 栃久保善博、橘内陽子、竹下潤一、渡邉美智子、佐々木崇光、保坂卓臣、吉成浩一　分子記述子を用いたリードアクロスによる反復投与毒性予測評価の検討　日本動物実験代替法学会　第 32 回大会（2019. 11）（つくば）
3) Tatsuya Ochibe, Kaori Ambe, Masahiro Tohkin: In silico models for predicting hepatotoxicity and renal toxicity based on HESS database. CBI 2019 (2019.10) (Funabori)
4) Yoko Kitsunai, Jun-ichi Takeshita, Michiko Watanabe, Takuomi Hosaka, Ryota Shizu, Takamitsu Sasaki, Kouichi Yoshinari: Read-across approach using molecular descriptors for the prediction of rat repeated-dose toxicity. EuroTox 2019, Helsinki, Finland, September, 2019.
5) Tatsuya Ochibe, Kaori Ambe, Masahiro Tohkin: In silico models for predicting of the repeated dose toxicity based on HESS database. ICT XV 2019, Honolulu, Hawaii, July, 2019.
6) Tatsuya Ochibe, Kaori Ambe, Masahiro Tohkin 機械学習法を利用した化学物質誘発性腎毒性の予測　第 1 回医薬品毒性機序研究会　(2019.1) (Nagoya)
7) 橘内陽子、竹下潤一、渡邉美智子、佐々木崇光、保坂卓臣、吉成浩一　計算科学に基づくリードアクロスによる反復投与毒性予測評価手法の開発　日本動物実験代替法学会　第 31 回大会　（2018. 10）（熊本）

３．The number and summary of patents and patent applications

It’s not applicable.

４．Others　(awards, press releases, software and database construction)

1) NCU Contact Points in Asia Symposium 2019 Best poster award (2019.12)(Nagoya) : Tatsuya Ochibe、Kaori Ambe、Masahiro Tohkin: Machine learning models for predicting hepatotoxicity and renal toxicity based on HESS toxicological test database.
2) 日本動物実験代替法学会　第32回大会　学生優秀演題賞受賞（2019.11、つくば）：栃久保善博、橘内陽子、竹下潤一、渡邉美智子、佐々木崇光、保坂卓臣、吉成浩一、分子記述子を用いたリードアクロスによる反復投与毒性予測評価の検討