

安全性データベースと機械学習を活用した肝悪性腫瘍誘発薬物予測手法の
開発に関する研究

A Study on the Development of Prediction Approaches for
Hepatocarcinoma-Inducing Drugs Utilizing Safety Databases and Machine
Learning Techniques

平成 31 年度入学 黒崎宏太 (Kurosaki, Kota)

近年、膨大で複雑なデータの関係を解析することにより高度な予測を達成する技術として、Machine Learning (ML) が注目されている。ML では、データの背後に存在するパターンを学習することによってモデルの構築を達成する。ML は、学習に用いられる説明変数と目的変数の組み合わせ方によって、多様な課題の解決に適用することができる。創薬の分野においては、蓄積された実験データに ML を広く適用することにより、化合物の様々な性質を説明変数として、薬効や毒性、薬物動態パラメータといった様々な目的変数の効率的な最適化が試みられている。本研究では、薬物の副作用の予測に注目した。

重篤な副作用は医薬品の市場撤退の原因となり得るため、創薬における重要な要因である。安全な医薬品を効率的に開発するために、開発初期において活用できる副作用の簡便かつ迅速な予測手法の確立が待ち望まれている。副作用は治療標的のみならず、意図されていない多様な生化学的経路に及ぼす影響によっても惹起される。そのため、副作用の理解と予測的評価には原因薬物に関する様々な生理活性情報の活用が必要となる。しかし、開発段階にある多くの候補化合物の網羅的な生理活性の評価には時間とコストを要するために、副作用に関連する生理活性情報の入手は困難であることが多い。

さらに、臨床試験における薬物の安全性情報は限定的であるために、予測のエンドポイントとなる副作用情報の収集も困難である。以上の様に、薬物の生理活性情報に基づいた副作用の理解と予測モデルの開発には、生理活性データと副作用データの不足が大きな問題となっていた。

予測モデルを開発するためのデータ不足に際し、既存の実験データや実臨床データを活用した新たな知見の創出は有望なアプローチとなり得る。前者の試みとして、これまでに、*in vitro* 試験データから化学構造情報と生理活性との関係をモデル化し予測する定量的構造活性相関(QSAR)解析を実施してきた¹⁾。QSAR解析によって構築される予測モデルは、被験薬物の欠落した生理活性情報を効率的に補完するための有用な技術になると考えられる。更に、後者の試みにおいて、近年の副作用研究では、大規模な医薬品副作用自発報告データベースを用いた新たな知見の創出が注目されている。本データベースは、単一施設では収集できない膨大な副作用情報を集積している。このような多量の実臨床データは、臨床試験では評価が困難であった副作用情報の解析に有用であると考えられる。従って、これらの分析手法の導入は、副作用の理解と予測モデルの開発において不足していた生理活性情報や副作用情報の検討に応用が可能であると考えられる。

本研究は、既存のデータから創出された知見を活用した副作用予測手法の開発を目的とした。予測のエンドポイントとして、薬物の市場撤退の最たる要因である肝毒性に着目した。中でも、事例が稀であるためにこれまで副作用情報の収集が困難であった肝悪性腫瘍の予測を志向し、*in vitro* 活性情報に基づく副作用予測モデルの開発を試みた。まず、化合物の毒性に関連する様々な生理活性情報の欠落を克服するために、既存の High-Throughput-Screening (HTS) データベースの活用によって、化学構造情報から網羅的な毒性標的に対するアゴニスト・アンタゴニスト活性を予測する ML モデルを

開発した。次に、有害事象自発報告データベースを活用することで、大規模な副作用報告から肝悪性腫瘍の報告頻度が特異的に高い薬物を検出した。最後に、これらの研究成果から新たに得られた種々の情報を学習データとして活用することにより、化学構造から推定された薬物の *in vitro* 活性に基づいて肝悪性腫瘍誘発薬物を予測する新たなフレームワークを開発した(Fig. 1)。

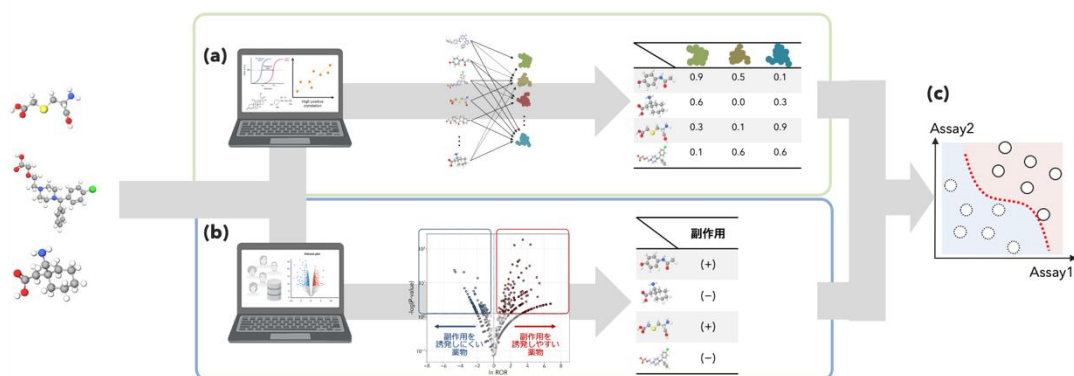


Fig.1 研究の概念図. (a) HTSデータベースを用いたQSARモデルの構築及び毒性作用の予測. (b) 有害事象自発報告データベースに基づく肝悪性腫瘍誘発薬物の検出. (c) 化学構造から推定された*in vitro* 活性に基づく肝悪性腫瘍誘発薬物の予測.

1. HTS データベースを活用した QSAR 予測ツールの開発

多くの化学物質は健康に与える影響が未知であるため、近年、QSAR モデルによる毒性評価の効率化が注目を集めている。このような背景から、米国では毒性データの収集・公開および新たな毒性評価手法の開発を志向し、大規模な毒性評価プロジェクトである Toxicology in the 21st Century Program (Tox21) が実施されている。Tox21 が提供する約 1 万化合物の HTS データベース(Tox21 データベース)には、毒性発現のハブとなる有害性経路である核内受容体やストレス応答経路に対する 59 種の *in vitro* 活性情報が格納されている。一般的に QSAR 解析を用いた既報の毒性予測モデルは予測ターゲットが限定的かつ非公開であり、QSAR モデルの網羅性と利便性に欠けている。そこで、本研究では Tox21 データベースを活用し、化学構造から 59 種の *in vitro* 活性を網羅的に予測する XGBoost モデルを開発した²⁾。開発されたモデルは良好な予測性能を示した(ROC 曲線下面積: 0.817 ± 0.088)。これ

らの結果より, QSAR モデルによる *in vitro* 活性の高精度予測が可能であると
考えられる. 本検討により構築された QSAR モデルは, Toxicity Predictor
と呼ばれるウェブアプリケーションツールとして一般公開された.

2. 副作用自発報告データベースを活用した肝悪性腫瘍発現傾向の解析

肝毒性の重篤な形態の 1 種である肝悪性腫瘍の報告が臨床において散見さ
れる. しかし, 臨床試験による安全性情報は, 母集団が小さく実施が短期間であるために, 限
定的である. 従って, 実臨床で報告された肝悪
性腫瘍に関連する薬物の知見は限られていた.
このような副作用に関する情報収集の限界か
ら, 有害事象自発報告データベースを用いた実臨床における安全性情報の活
用に着目した. これまでに, FDA Adverse Event Reporting System (FAERS)
に基づき, 有害事象と薬物の多様な関係が検討されてきた. そこで, 本研究
では FAERS に基づき肝悪性腫瘍の報告頻度が特異的に高い薬物の検出を試
みた³⁾. Fig. 2 は, FAERS から算出した報告オッズ比(ROR)と Fisher の正確
検定における P 値を用い, 肝悪性腫瘍に関連する薬物を示したものである.
図の右上は肝悪性腫瘍発現に有意に関連する ROR が 1 より大きい薬物を示
している. 解析の結果, 感染性肝疾患治療薬など, 肝悪性腫瘍のリスクが高
い背景を有する集団の治療薬が検出された. 本結果から, 背景疾患による交
絡が疑われたため, 多重ロジスティック回帰分析により, これらの背景を調
整した詳細な検討を行った. その結果, Quinine, Mianserin, Tolvaptan など
を代表とする計 47 種の薬物が検出され, 本薬物群を肝悪性腫瘍誘発薬物と
した.

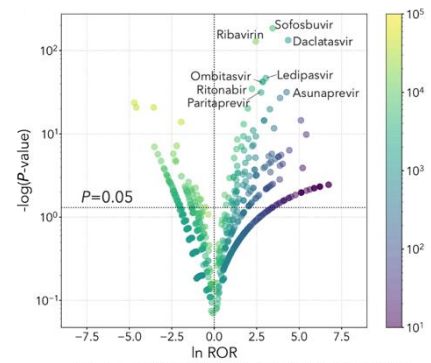


Fig2. 肝悪性腫瘍報告に関連する薬物

3. *In vitro* 活性情報に基づく肝悪性腫瘍誘発薬物の予測

創薬プロセスの短縮とコスト削減のため、毒性に対する様々な *in silico* 予測モデルが提案されている。しかし、臨床で観察される肝悪性腫瘍の薬理的予測因子については、副作用情報と薬物の生理活性データの欠落のために、知見が限られている。そこで、先に述べた *in vitro* 活性予測ツールおよび医薬品副作用自発報告データベースから得られた知見を

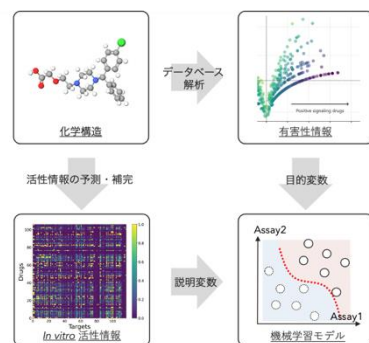


Fig. 3 QSARにより推定された *in vitro* 活性作用に基づく肝悪性腫瘍誘発薬物予測のフレームワーク

活用した新たな肝悪性腫瘍に関する予測モデル構築法を探求した⁴⁾。すなわち、第1章で構築した QSAR モデルによって予測された *in vitro* 活性情報を説明変数、FAERS から検出された肝悪性腫瘍誘発薬物を目的変数として活用し、MLモデルの学習を達成することで、肝悪性腫瘍誘発薬物を予測する新たなフレームワークを提案した(Fig. 3)。本スキームに基づき ML モデルを構築し、外部検証データを用いて ML モデルの汎化性能を評価した。その結果、肝悪性腫瘍誘発薬物の予測において、ROC 曲線下面積が 0.7 以上の高い性能を認めた。LightGBM モデルの解析の結果、予測に強く貢献した代表的な因子としてゲノム不安定性に関連する ATAD5、細胞増殖に関連するエストロゲン受容体 α (ER- α)、エストロゲン関連受容体(ER α)、甲状腺刺激ホルモン放出ホルモン受容体(TRHR)などが検出された。また、LightGBM により真陰性判定された

		Hepatocarcinoma	ATAD5 inducer	TRHR agonist	ER- α antagonist	ERR antagonist	ERR agonist
	(+)	Experimental Active	Predicted Active	Predicted Active	Predicted Active	Predicted Active	Predicted Active
	(+)	Experimental Active	Predicted Active	Predicted Active	Predicted Active	Predicted Active	Predicted Active
	(-)	Experimental Inactive	Predicted Inactive	Predicted Inactive	Predicted Inactive	Predicted Inactive	Predicted Inactive

Table 1 LightGBMから特定された重要な *in vitro* 活性情報

Mirtazapine に対し、真陽性判定された Mianserin や Tacrolimus では、これらの *in vitro* 活性が高いことが観察された(Table 1)。これらの結果は、

QSAR によって推定された *in vitro* 活性情報と FAERS から検出された安全性情報を組み合わせることで、生理活性情報に基づいた肝悪性腫瘍誘発薬物の予測とメカニズムの推定が可能であることを示している。

4. 総括

本研究では、肝悪性腫瘍の副作用予測を志向し、QSAR モデルならびに有害事象自発報告データベースの新たな活用法を検討した。副作用予測の試みは、エンドポイントの評価が困難であることに加え、副作用の発現に複数の生理機能が介在する可能性がある。このような背景から、予測手法の確立とその背後に存在する薬物標的の理解は困難な課題であった。本研究では、ML 手法を活用することにより、薬物の欠落した *in vitro* 活性情報を高度に予測補完する QSAR モデルを構築したことに加え、自発報告データベースから安全性を評価することにより、稀な副作用である肝悪性腫瘍を予測対象とした新たな学習データの構築を達成した。さらに、これらの成果を活用することで、QSAR 解析により推定された生理活性情報から誘発機序の理解と副作用予測が可能であることを示した。本研究成果が、安全性データの欠落を克服し、創薬における毒性評価を効率化するための新たなアプローチとなることを期待する。

《参考文献》

- 1) Yamali C., Sakagami H., Uesawa Y., Kurosaki K., Satoh K., Masuda Y., Yokose S., Ece A., Bua S., Angeli A., Supuran C.T., Gul H.I., *Eur. J. Med. Chem.*, **217**, 13351 (2021).
- 2) Kurosaki K., Raymond W., Uesawa Y., *Int. J. Mol. Sci.*, **21**(21), 7853 (2020).
- 3) Kurosaki K. and Uesawa Y., *Biomolecules*, **11**(7) 944 (2021).
- 4) Kurosaki K. and Uesawa Y., *J. Toxicol. Sci.* **47**(3) 89-98 (2022).