

京都大学	博士（薬科学）	氏名	酒井 幸
論文題目	化学構造式から受容体親和性を予測する機械学習モデルの構築とヒト有害事象の予測への応用		

(論文内容の要旨)

多くの医薬品は簡単な化学構造式で表現可能な化合物であるが、経験を積んだ研究者でも化学構造式をみただけで作用を予測することは難しい。しかし薬理作用の程度や様式を決定するのはその化学構造であるから、機械学習等で情報を適切に抽出できれば、原理的には化学構造から薬理作用が予測可能と考えられる。医薬品はその治療標的となる生体分子に強く作用するが、程度の差こそあれ他の生体分子にも作用する。そのため安全性試験を通過した医薬品でさえも、市販後に想定外の有害事象が発現し、臨床上的問題となる。そこで本研究では、第1章で化学構造式を入力として種々の標的タンパク質（受容体）に対する親和性予測モデルを構築した。その上で第2章では親和性予測値とヒトでの有害事象データを関連づけ、既存の医薬品のみならず作用未知の化合物の薬効や有害事象の予測を化学構造式のみから可能とするプラットフォームを構築した。

第1章 グラフ畳み込みニューラルネットワークを用いた化学構造式からの親和性予測

親和性予測モデル構築に用いる化学構造式は ChEMBL(化合物の構造と受容体親和性実測値の公開データベース)より抽出した。受容体としては8種のタンパク質ファミリーに属す542種の多様な生体分子を選択した。化学構造の深層学習はグラフ畳み込みニューラルネットワーク(GCN)の手法を用いた。高精度な定量的予測モデルを得るために、GCNの隠れ層サイズを従来法である定性的予測モデルのそれと比べて9-32倍拡大し、複数モデルのアンサンブル使用により予測性能の向上を達成した。予測性能の目安となる平均絶対誤差値(pIC₅₀の実測値と予測値の差)は542種中395種(73%)で0.6以下、521種(96%)で0.7以下であった。予測モデルの妥当性は2種類の方法で検証した。まず画像認識等で頻用される畳み込み計算手法による親和性予測モデルと比較したところ、従来法の二乗平均平方誤差が0.68であるのに対して、提案手法は0.49とより高精度での予測が可能であった。次に抗うつ薬の作用点であるセロトニントランスポーター(SERT)に対する作用がまだ調べられていない化合物をChEMBLから抽出してSERT阻害活性を予測した。阻害活性予測値(IC₅₀)が10nMであったCHEMBL1377753の合成品を入手してSERT発現細胞での阻害活性を実測したところIC₅₀は6.24nMで予測に近似しており、既存抗うつ薬と同程度の強さであった。また当該化合物の*in vivo*作用を代表的な抗うつ薬評価系であるマウス尾懸垂試験で実測したところ、有意な抗うつ作用が認められた。以上より、多様な生体分子に対して高い精度と予測妥当性を併せ持つ親和性予測モデルの構築に成功した。

第2章 受容体親和性予測値を用いた有害事象および原因標的タンパク質の予測

近年、市販後に明らかになった有害事象(Adverse Event, AE)ビッグデータの解析により、AEの発現に対する多剤併用の影響や、併用時のAE低減傾向に着目した新規創薬標的の提案が報告されている。しかし薬理活性に基づくAE予測

の研究例は数少ない。本研究では、AE 発現前に使用していた全ての医薬品がその AE に関与しえると仮定し、全医薬品成分の 542 種の受容体に対する親和性予測値と AE 報告数との関係性に基づく AE 発症予測を試みることで、AE の発生原因となりえる受容体を明らかにしようとした。我が国の医薬品副作用報告データを用いて SMQ と呼ばれる 166 種の AE グループに対して予測モデルを構築したところ、AE 報告数が 100 件以上存在する 98 種の SMQ のうち 72 種(74%)において予測モデルの AUC (予測性能指標) が 0.7 以上の精度で予測が可能となった。これによってトルサード・ド・ポアント/QT 延長不整脈に対する hERG K⁺チャネルや内皮型一酸化窒素合成酵素など、AE 発症への関与が疑われる標的タンパク質が 69 種(70%)の SMQ において浮かび上がった。関与が疑われる標的タンパク質としての報告の少ない生体分子も数多く含まれていた。また、動物による検証が不可能な、例えば自殺・自傷関連有害事象についても関連が疑われる新規標的タンパク質の存在が示唆された。

以上、申請者は、化学構造式から生体分子への親和性を高精度に予測する学習モデルを確立し、その親和性予測値によりヒトでの薬効・有害事象が予測可能であることを示した。本予測モデルは既存の医薬品のリポジショニングや作用未知の化合物のヒトでの薬理作用を予測するプラットフォームとしての活用が期待される。

※ 学位授与された方の「論文内容の要旨」、「論文審査結果の要旨」(審査教員作成)は、学位授与日から 3 ヶ月以内に京都大学学術情報リポジトリに掲載され公開されます。学位申請を行う方は掲載を承認されたものとします。

(論文審査の結果の要旨)

多くの医薬品は簡単な化学構造式で表現可能な化合物であるが、経験を積んだ研究者でも化学構造式をみただけで作用を予測することは難しい。しかし薬理作用の程度や様式を決定するのはその化学構造であるから、機械学習等で情報を適切に抽出できれば、原理的には化学構造から薬理作用が予測可能と考えられる。医薬品はその治療標的となる生体分子に強く作用するが、程度の差こそあれ他の生体分子にも作用する。そのため安全性試験を通過した医薬品でさえも、市販後に想定外の有害事象が発現し、臨床上の問題となる。そこで本研究では、第1章で化学構造式を入力として種々の標的タンパク質(受容体)に対する親和性予測モデルを構築した。その上で第2章では親和性予測値とヒトでの有害事象データを関連づけ、既存の医薬品のみならず作用未知の化合物の薬効や有害事象の予測を化学構造式のみから可能とするプラットフォームを構築した。

第1章 グラフ畳み込みニューラルネットワークを用いた化学構造式からの親和性予測

親和性予測モデル構築に用いる化学構造式は ChEMBL(化合物の構造と受容体親和性実測値の公開データベース)より抽出した。受容体としては8種のタンパク質ファミリーに属す542種の多様な生体分子を選択した。化学構造の深層学習はグラフ畳み込みニューラルネットワーク(GCN)の手法を用いた。高精度な定量的予測モデルを得るために、GCNの隠れ層サイズを従来法である定性的予測モデルのそれと比べて9-32倍拡大し、複数モデルのアンサンブル使用により予測性能の向上を達成した。予測性能の目安となる平均絶対誤差値(pIC₅₀の実測値と予測値の差)は542種中395種(73%)で0.6以下、521種(96%)で0.7以下であった。予測モデルの妥当性は2種類の方法で検証した。まず画像認識等で頻用される畳み込み計算手法による親和性予測モデルと比較したところ、従来法の二乗平均平方誤差が0.68であるのに対して、提案手法は0.49とより高精度での予測が可能であった。次に抗うつ薬の作用点であるセロトニントランスポーター(SERT)に対する作用がまだ調べられていない化合物を ChEMBL から抽出して SERT 阻害活性を予測した。阻害活性予測値(IC₅₀)が10 nMであった ChEMBL1377753 の合成品を入手して SERT 発現細胞での阻害活性を実測したところ IC₅₀ は6.24 nM で予測に近似しており、既存抗うつ薬と同程度の強さであった。また当該化合物の *in vivo* 作用を代表的な抗うつ薬評価系であるマウス尾懸垂試験で実測したところ、有意な抗うつ作用が認められた。以上より、多様な生体分子に対して高い精度と予測妥当性を併せ持つ親和性予測モデルの構築に成功した。

第2章 受容体親和性予測値を用いた有害事象および原因標的タンパク質の予測

近年、市販後に明らかになった有害事象(Adverse Event, AE)ビッグデータの解析により、AEの発現に対する多剤併用の影響や、併用時のAE低減傾向に着目した新規創薬標的の提案が報告されている。しかし薬理活性に基づくAE予測の研究例は数少ない。本研究では、AE発現前に使用していた全ての医薬品がそのAEに関与しえると仮定し、全医薬品成分の542種の受容体に対する親和性予測値とAE報告数との関係性に基づくAE発症予測を試みることで、AEの発生原因となりえる受容体を明らかにしようとした。我が国の医薬品副作用報告データ

を用いて SMQ と呼ばれる 166 種の AE グループに対して予測モデルを構築したところ、AE 報告数が 100 件以上存在する 98 種の SMQ のうち 72 種(74%)において予測モデルの AUC (予測性能指標) が 0.7 以上の精度で予測が可能となった。これによってトルサード・ド・ポアント/QT 延長不整脈に対する hERG K⁺チャネルや内皮型一酸化窒素合成酵素など、AE 発症への関与が疑われる標的タンパク質が 69 種(70%)の SMQ において浮かび上がった。関与が疑われる標的タンパク質としての報告の少ない生体分子も数多く含まれていた。また、動物による検証が不可能な、例えば自殺・自傷関連有害事象についても関与が疑われる新規標的タンパク質の存在が示唆された。

以上、申請者は、化学構造式から生体分子への親和性を高精度に予測する学習モデルを確立し、その親和性予測値によりヒトでの薬効・有害事象が予測可能であることを示した。本予測モデルは既存の医薬品のリポジショニングや作用未知の化合物のヒトでの薬理作用を予測するプラットフォームとしての活用が期待される。

よって、本論文は博士(薬科学)の学位論文として価値あるものと認める。また、令和 4 年 2 月 17 日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。

要旨公表可能日： 年 月 日以降