

和漢薬のターゲットタンパク質のインシリコ探索

統括者	梅壽 雅人	情報科学分野	特命准教授
所内共同研究者	門脇 真	消化管生理学分野	教授
申請代表者	立川 仁典	横浜市立大学大学院生命ナノシステム科学研究科	教授
所外共同研究者	山西 芳裕	九州大学高等研究院	准教授
所外共同研究者	五斗 進	京都大学化学研究所バイオインフォマティクスセンター化学生命科学研究領域	准教授

【報告セミナー要旨】

【要旨】

和漢薬の化学的学術情報と臨床効果に関する情報は蓄積されてきたが、その作用機序の科学的根拠は未だ乏しい。本研究では、複雑系である和漢薬の作用機序解明を目指し、既知の和漢薬標準化合物に関するケモインフォマティクス分野の情報と、ヒトゲノム解析などのバイオインフォマティクス分野の情報から有益な情報を抽出し、和漢薬標準化合物の標的タンパク質探索を目的とするインシリコ解析を実施する。和漢医薬学総合研究所の探索研究プロジェクトにおいて提供されている和漢薬標準化合物（探索研究プロジェクト：96種類）のヒト全タンパク質に対するインシリコでの結合シミュレーション解析や和漢薬成分の標的タンパク質予測のための機械学習法を開発し、和漢薬の標的タンパク質を解析するためのデータベースを確立する。

和漢薬標準化合物に関して、ヒト全タンパク質立体構造に対する結合シミュレーション（約400万パターン）は、ほぼ計算が終了しその結果（結合エネルギー、結合状態の分子イメージなど）を表示するデータベースを構築している。また、機械学習手法による予測手法確立のために、上記の和漢薬標準化合物に関して、最新の化学構造記述子に基づく化学構造類似度を計算した。少なくとも一つ以上のタンパク質との相互作用情報を持つ約50万個の化合物に対して類似度検索を行い、和漢薬成分とタンパク質のペアの候補を6000件以上挙げた。その結果をデータベースとしてまとめている。また、試作公開内部版（富山大学杉谷キャンパス内のコンピュータからのみ閲覧でき、IDとパスワードで保護している）として、予測された標的タンパク質と類似化合物の「結合強度情報」や、そのタンパク質が関連する「疾病情報」に関する情報を閲覧できるようにしている（これらの解析手法には特許に絡む事項があるため公開を内部に限定している）。6,000件以上の予測標的タンパク質に関する情報のうち、「結合強度情報」が提示されているデータが約1,700件以上あり、「疾病情報」が提示されているデータが約1,300件以上存在している。「結合強度情報」や「疾病情報」がない類似化合物データについても、全ての化合物に「相互作用する標的タンパク質の情報や、そのタンパク質のオントロジー情報、パスウェイ情報」などが付加されており、今後の解析で和漢薬成分の作用機序の解明に有用なデータを抽出する。

現在、これらの結果を元にオントロジー解析や、パスウェイ解析の方法を開発しており、結合シミュレーションと機械学習法の結果をリンクさせる機能も構築している。今後、オントロジー解析やパスウェイ解析などの可視化及び、結合シミュレーションとの連携を進めて、和漢薬標準化合物の標的タンパク質候補を予測する。

最終的には、複合薬物である漢方薬の作用機序の解明を目指して、漢方薬に含まれる多数の化合物とそれらの標的となる多数の標的タンパク質の相互作用の解明を目標とした『「多成分」対「複合ターゲット」』という複雑系解析手法を開発する。

■背景および目的

和漢薬の化学的学術情報と臨床効果に関する情報は蓄積されてきたが、その作用機序の科学的根拠は未だ乏しい。そこで、複雑系である和漢薬の作用機序解明を目指し、既知の和漢薬標準化合物に関するケモインフォマティクス分野の情報と、ヒトゲノム解析などのバイオインフォマティクス分野の情報から有益な情報を抽出し、和漢薬標準化合物の標的タンパク質探索を目的とするインシリコ解析を実施する。具体的には、本課題が採択されている公募型研究において提供されている和漢薬標準化合物（探索研究プロジェクト：96種類）のヒト全タンパク質に対するインシリコでの結合シミュレーション解析を実施すると共に、既知医薬品等の化合物と標的タンパク質間の相互作用情報から化学構造類似度に基づいた和漢薬成分の標的タンパク質の予測システムを構築し、和漢薬の標的タンパク質を解析する手法を確立する。

本年度は、膨大な化合物の化学構造や薬理情報を表すケミカル空間とヒトのタンパク質をコードする全遺伝子の様々な情報を表すゲノム空間の関係を探索するケモゲノミクスのインシリコ手法を適用し、和漢薬標準化合物の標的タンパク質の大規模探索を行う。そのため、既知の化合物・タンパク質間相互作用のインターラクトーム情報を様々なデータベースや文献情報から抽出し、ケモゲノミクスの予測モデルを学習するための大規模な化合物・タンパク質間相互作用の学習データを構築する。それに基づき学習した予測モデルを、本研究で対象とする和漢薬標準化合物や生薬含有化合物に対して適用し、それらの標的タンパク質やオフ標的タンパク質（本来目標としていない標的タンパク質）をゲノムワイドに予測する。これらの予測結果や解析結果を系統的に整備し、データベースとして公開する。

上記の和漢薬標準化合物の標的タンパク質解析では、『「1成分」対「複合ターゲット」』という複雑系の解析となるが、最終的には『「多成分」対「複合ターゲット」』である和漢薬の複雑な作用メカニズムの解明を目指して、生薬に含まれる多数の化合物と、それらの標的となる多数の標的タンパク質の相互作用解明を目標とした複雑系解析手法の開発を目的とする。

■方 法

2-1. 和漢薬成分の標的タンパク質の予測システム開発：インシリコ結合シミュレーションによる予測

上記、96種類の和漢薬標準化合物に関して、ヒト全タンパク質立体構造（約4万種類）に対する結合シミュレーション（約400万件）を実施し、標的タンパク質及びその作用部位を予測する。

2-2. 生薬を構成する多成分系化合物情報の整備

京都大学化学研究所が開発しているPCIDB（<http://www.genome.jp/db/pcidb>）のデータを利用して、和漢研・拠点推進室が配布している120種類の生薬について、その代謝産物である約2700種類の生薬含有多成分系化合物の分子構造情報を取得する。

2-2. 和漢薬成分の標的タンパク質の予測システム開発：機械学習手法による予測

ChEMBL、BindingDBなどの公共データベースや文献情報から、化合物-タンパク質相互作用情報を抽出する。ノイズを除去し、相互作用の結合強度の強い化合物タンパク質ペアだけで構成されたデータを整備する。最新の化学構造記述子KCF-S (KEGG Chemical Function Substructure)に基づく化学構造類似度を計算する。化合物の化学構造類似性に基づく教師付き学習の枠組みで、96種類の和漢薬標準化合物と約2700種類の生薬成分化合物の標的タンパク質を予測する機械学習の手法を開発する。

2-3. 和漢薬成分の作用機序の解析や適応可能疾患の予測

次に、上記方法で漢方薬成分が相互作用すると予測された標的タンパク質について、タンパク質の生物学的機能で分類するオントロジー解析 (Gene OntologyやKEGG Orthologyなど) を行い、タンパク質の機能グループを同定する。さらに予測された標的タンパク質群を既知のパスウェイ (KEGGやReactomeなどのシグナリング経路や代謝経路) にマッピングし、影響を及ぼしている作用パスウェイなどを同定する。標的タンパク質や作用パスウェイの結果を治療標的の点から考察し、和漢薬成分の適応可能疾患を予測する。

2-4. 和漢薬の作用機序の解析や適応可能疾患の予測

上記の標的タンパク質予測、オントロジー解析、パスウェイ解析、適応可能疾患予測を、生薬のレベルと方剤のレベルで階層的に行う。これによって、和漢薬が作用する標的タンパク質の機能グループ、作用パスウェイ、適応可能疾患をそれぞれ予測する。

■結果および考察

96種類の和漢薬標準化合物に関して、ヒト全タンパク質立体構造に対する結合シミュレーション (約400万パターン) は、ほぼ計算が終了しその結果 (結合エネルギー、結合状態の分子イメージなど) を表示するデータベースを構築した (試作公開版: <http://160.26.6.205/ligand/wakan/koba3.jsp>, 図1)。

機械学習手法による予測手法確立のために、約120万個の化合物・タンパク質相互作用ペアを整備し、これを学習データとして機械学習の予測モデルを構築した。その予測モデルを上記96種類の和漢薬標準化合物と2700種類の生薬含有化合物に適用し、和漢薬成分とタンパク質のペアの相互作用候補を列挙した。

予測された和漢薬成分の標的タンパク質に対して、オントロジー解析、パスウェイ解析、適応可能疾患予測を行った。これによって和漢薬の各成分が作用する可能性のある標的タンパク質の機能グループと作用パスウェイを同定し、治療標的の観点から各和漢薬成分の適応可能疾患を新規予測した。今後の解析で和漢薬成分の作用機序の解明に有用なりソースとなることが期待できる。

予測された和漢薬成分の標的タンパク質情報をマージし、248個の和漢薬に対して標的タンパク質予測、オントロジー解析、パスウェイ解析、適応可能疾患予測の行程を、生薬のレベルと方剤のレベルで階層的に行った。これによって248個の和漢薬が作用する可能性のある標的タンパク質の機能グループと作用パスウェイを同定し、治療標的の観点から和漢薬の適応可能疾患を新規予測した。これらの解析で得られた知見は、実際の和漢薬を臨床現場で有効活用するのに有用なりソースとなること

が期待できる。

96種類の和漢薬標準化合物に関する予測結果や解析結果はデータベースとしてまとめている（試作公開版：http://160.26.6.205/ligand/wakan/chembl_4/、図2）、生薬含有化合物に関しても近日中にデータベースに取り込む予定である。また、試作公開内部版（富山大学杉谷キャンパス内のコンピュータからのみ閲覧でき、IDとパスワードで保護している）として、タンパク質が治療標的として関連する疾病情報を閲覧できるようにしている（これらの解析手法には特許に絡む事項があるため公開を内部に限定している）。

■結論

特定研究課題「和漢薬のターゲットタンパク質のインシリコ探索」の研究として、探索研究プロジェクトで提供している96種類の和漢薬標準化合物について、ヒト全タンパク質との結合シミュレーション（約400万件）で標的タンパク質や作用部位を予測した。化合物タンパク質間相互作用インターラクトーム情報に基づく機械学習によって、上記96種類の和漢薬標準化合物と約2700種類の生薬含有化合物の標的タンパク質を大規模予測した。標的タンパク質の機能グループや作用パスウェイの同定し、適応可能疾患の新規予測を行った。これらの結果を整備し、その一部に関してはデータベース化を行った。

昨年度に実施した和漢薬成分の標的タンパク質解析では、『「1成分」対「複合ターゲット」』という複雑系の解析手法を開発したが、本年度は『「多成分」対「複合ターゲット」』である和漢薬の複雑なメカニズムの解明を目指して、生薬に含まれる多数の化合物と、それらの標的となる多数の標的タンパク質の複雑系解析手法を確立することができた。今後、全ての解析結果を整備して、データベース化する予定である。

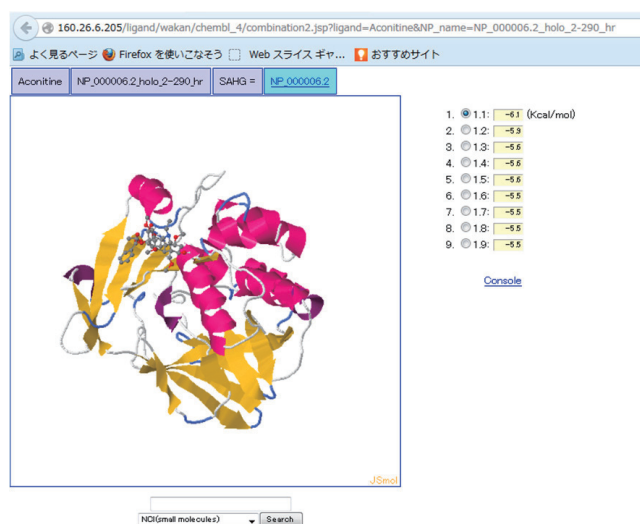


図1、和漢薬標準化合物とヒト全タンパク質との結合シミュレーション結果表示。

この結合シミュレーション結果の表示には「JsMol」を利用しているので、従来の「Jmol」を使った場合のセキュリティー懸念による非表示という問題はない。タンパク質は分子表面で表示され、化合物はボール&スティックで表示される。結合状態はエネルギー順位のラジオボタンで選択することができ、全ての立体構造は「JsMol」による3次元表示や拡大・縮小などの操作を利用できる。

