

薬剤標的タンパク質と副作用の関連分析のための統計手法

A statistical framework for analyzing the association between drug target proteins and drug side-effects

化学研究所 化学生命科学 水谷紗弥佳

背景と目的

多くの薬剤は特定のタンパク質を標的とし、細胞内のタンパク質ネットワーク、あるいは分子経路全体の制御を誘導する小化合物である。薬剤の投与は、期待通りの治療効果をもたらす一方、望ましくない副作用を引き起こすこともある。副作用は軽少なものから重篤なものまであり、医薬品開発において重要な懸案であるにも関わらず、その分子作用機構はほとんど未同定のままである。近年、医薬品関連情報は公的に利用可能なデータベースに大量に蓄積されてきた。そのような情報は、薬剤の化合物構造、標的タンパク質、薬効、副作用など様々なデータを含み、その間の関係を計算器学的アプローチによって明らかにすることが期待されている。この研究では、薬剤-標的タンパク質の相互作用という分子レベルでの情報と、薬剤の副作用という表現型レベルでの情報という、異なるレベルでの情報を関連づける方法を提案する。

検討内容

薬と標的タンパク質の相互作用情報を DrugBank と Matador データベースから取得し、相互作用の有無により二値行列を作成。一方、同じ薬に対して副作用との相互作用情報を SIDER データベースから取得し、同じように二値行列を作成。スパース標準相関解析 (sparse canonical correlation analysis) を適用し、2つの行列の間の相関を最大にするようなタンパク質と副作用の組み合わせを求めた。このようなタンパク質副作用の組み合わせを相関セットと呼ぶ。

結果

スパース標準相関解析を薬剤-標的タンパク質相互作用行列と薬剤-副作用行列に適用し、2つの行列の間に相関のあるタンパク質と副作用の相関セットを抽出した (図 1)。

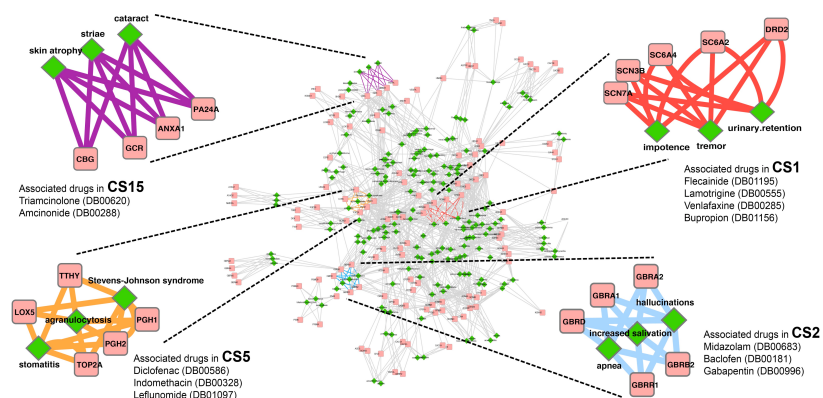


図 1 スパース標準相関解析により同じ相関セット (CS) に抽出されたタンパク質 (オレンジ四角) と副作用 (緑ダイヤ) のネットワーク

抽出された相関セットの生物学的妥当性は、Kyoto Encyclopedia of Genes and Genomes (KEGG) と Gene Ontology (GO) データベースを用いた、タンパク質分子経路及びタンパク質機能分類による enrichment 解析で評価した (図 2)。その結果、同じ分子経路で働くタンパク質は、分子機能が異なっているにもかかわらず同じ相関セットにクラスタリングされたことが分かった。すなわち、ある相関セットに抽出されたタンパク質群の薬剤による機能制御は、それらが異なるタンパク質機能分類に属していても同じ分子経路の活性化、不活性化を通じて、同じような副作用を引き起こすのではないかと解釈することが出来る。さらに、スパース標準相関解析を薬剤の標的タンパク質プロファイルに適用し、その副作用を予測する応用を考えた。その予測精度を検証したところ、先行研究により提唱された方法より上回る予測精度を得た (表 1)。

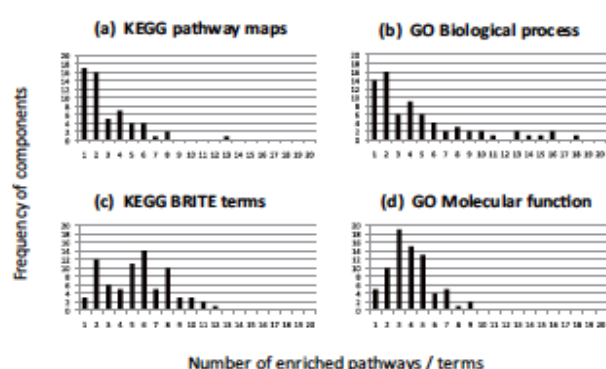


図 2 パスウェイ分類 (上 2 つ) と機能分類 (下 2 つ) による相関セットの分布。(a, b) KEGG による分類; (b, d) GO による分類

考察

今回提案した手法は、副作用の分子作用機序への議論を可能にするとともに、薬剤候補化合物に対して標的タンパク質プロファイルから潜在的な副作用を予測するのに有用であると期待される。今後の課題としては、それらのタンパク質分子経路の薬剤による制御が副作用を起こすメカニズムを推定することである。また、データベースに登録されている副作用には、純粋に薬剤によるものと、患者側の既往症や体質によるものがあるが、それをある程度分けて考える事が重要である。

参考文献

Mizutani, S., Pauwels, E., Stoven, V., Goto, S., and Yamanishi, Y.; Relating drug-protein interaction network with drug side-effects. *Bioinformatics* 28, i522-i528 (2012).

表 1 予測精度の比較

Method	AUC $\pm$ S.D.	AUPR $\pm$ S.D.
Chemical structure:		
Random	0.5000 $\pm$ 0.0000	0.0556 $\pm$ 0.0000
OCCA	0.8355 $\pm$ 0.0010	0.3753 $\pm$ 0.0016
SCCA	0.8708 $\pm$ 0.0007	0.3766 $\pm$ 0.0030
Target protein:		
Random	0.5000 $\pm$ 0.0000	0.0556 $\pm$ 0.0000
OCCA	0.8850 $\pm$ 0.0007	0.4067 $\pm$ 0.0006
SCCA	0.8895 $\pm$ 0.0002	0.4103 $\pm$ 0.0018

(太文字) 今回提案した手法が与えた予測精度。(その他) 先行研究により提唱された方法の予測精度。