

疾患と DrugMatrix データセット との間の遺伝子発現の 統合解析におけるテンソル分解を用いた 教師なし学習による変数選択を用いた候補薬物の同定

田口 善弘^{1,a)}

概要: 創薬はお金と時間がかかるプロセスであり、わずかでもその時間と費用を削減するため、インシリコ創薬に大きな期待が寄せられている。インシリコ創薬の二大潮流である構造ベースとリガンドベースの創薬には長所と短所があり、新たな方法として近年、遺伝子発現プロファイルを用いる研究が行われるようになった。しかし、その多くは教師あり学習であり、既知の薬物が知られていない疾患・標的蛋白についてのインシリコ創薬を行うのは難しい。本研究では著者がかねてから提案しているテンソル分解を用いた教師なし学習による変数選択を用いたインシリコ創薬を提案し、既知薬剤がない場合でも遺伝子発現プロファイルから創薬できるようにすることを目指す。

Y-H. TAGUCHI^{1,a)}

1. はじめに

創薬においてインシリコ創薬を用いることでその費用と時間を短縮しようという試みがなされるようになってから久しい [2]。インシリコ創薬には二大潮流がありそれはタンパクの構造に基づく創薬 (Structure Based Drug Design; SBDD) と化合物の構造に基づく創薬 (Ligand based Drug Design; LBDD) である。SBDD はタンパクの立体構造を元に、化合物とのいわゆるドッキングシミュレーションを行うことで創薬を行うため、いかなる場合でも創薬候補化合物を提示することが可能であり、低分子化合物に限らず、中分子や、ペプチド創薬に適用可能であるなどの長所を持つ。一方で、標的タンパクの構造が未知の場合にはまず構造推定を行わなくてはならないため、そこで予測精度が下がってしまう、非常に計算負荷が高いため、数百万個の化合物ライブラリにそのまま適用するのは難しい、などの短所がある。一方、LBDD は化合物の情報だけから新規化合物を探索するため、計算負荷も軽く、また、予測精度も高いという利点を持つが、一方で、既知の薬剤化合物がない場合、無力である、既知の薬剤化合物と構造的に

類似性が低い新規薬剤化合物の探索は難しい、などの欠点もある。

このような SBDD, LBDD の欠点を補うため、遺伝子発現プロファイルからのインシリコ創薬が試みられるようになった [3]。遺伝子発現プロファイルからのインシリコ創薬では、各種化合物を培養細胞やモデル動物に投与した際の遺伝子発現プロファイルの類似性を検索し、似た発現プロファイルを持つものは同じ標的タンパク・疾患を共有しているであろうという仮定の元に新規薬剤候補化合物を探索する。この方法は、遺伝子発現プロファイルの計測という新たなコストが発生するものの、その費用は創薬に要する費用に比べれば僅かであり、また、一度計測してしまえば、再度計測する必要のない化合物の特徴量として利用可能であるため、その費用対効果は高いと期待される。

このような長所をもつ遺伝子発現プロファイルベースのインシリコ創薬であるが、LBDD と同じように、既知の新規薬剤化合物がない場合には、新規薬剤候補化合物を同定できないという、LBDD と同じ教師あり学習における限界からは逃れようがない。本稿では遺伝子発現プロファイルベースのインシリコ創薬を、かねてから著者が提唱しているテンソル分解を用いた教師なし学習による変数選択 [4-10] を用いて、既知の薬剤化合物なしでも行えるように拡張することを目指す。

¹ 中央大学理工学部物理学科
東京都, 112-8551, 日本

^{a)} tag@granular.com

本研究は原著論文として出版済みである。[1]

2. データと手法

2.1 データ

本研究には、化合物を投与した際の遺伝子発現プロファイルと、対象疾患の遺伝子発現プロファイルの二種類が必要である。前者については DrugMatrix(GEO ID:GSE59927) を用いた。DrugMatrix はラットに多種の化合物を投与したときの多種の臓器における遺伝子発現プロファイルの、投与後の時間変化を網羅的に記録した大規模なデータベースである。後者については GEO を検索し、疾患ごとに適当なものを探した。具体的には以下の通り。1) 心臓疾患: DrugMatrix から化合物投与時のラットの心臓の発現プロファイル (GSE59905) を、疾患側からは人の心臓疾患の遺伝子発現プロファイル (GSE57345) を用いた。2) 心的外傷後ストレス障害 (PTSD): DrugMatrix から化合物投与時のラットの脳の発現プロファイル (GSE59895) を、疾患側からは PTSD モデルラットの海馬と扁桃体の遺伝子発現プロファイル (GSE60304) を用いた。3) 急性リンパ性白血病 (ALL): DrugMatrix 側からは化合物投与時のラットの骨髄の遺伝子発現プロファイル (GSE59894) を、疾患側からは ALL 患者の骨髄の遺伝子発現プロファイル (GSE67684) を用いた。4) 糖尿病: DrugMatrix 側からは化合物投与時のラットの腎臓の発現プロファイル (GSE59913) を、疾患側からは人間の糖尿病患者の腎臓の遺伝子発現プロファイル (GSE30122) を用いた。5) 腎臓がん: DrugMatrix 側からは化合物投与時のラットの腎臓の発現プロファイル (GSE59913) を、疾患側からは人間の腎臓がん患者の腎臓の遺伝子発現プロファイル (GSE40435) を用いた。6) 肝硬変: DrugMatrix 側からは化合物投与時のラットの肝臓の遺伝子発現プロファイル (GSE59923) を、疾患側からは詳しくは肝硬変の患者の肝臓の遺伝子発現プロファイル (GSE15654) を用いた。詳しくは原著論文 [1] を参照されたい。

2.2 テンソル分解

まず、テンソル分解について述べる。テンソル分解にはいろいろな種類があるが [11]、本稿ではタッカー分解と呼ばれる分解を用いる。簡単のため三相のテンソルを用いて説明するが、四相以上の場合への拡張は自明であろう。要素が $x_{ijk} \in \mathbb{R}^{N \times M \times K}$ であるような三相のテンソル $\mathcal{X}$ を考える。この時、タッカー分解は

$$x_{ijk} = \sum_{\ell_1=1}^N \sum_{\ell_2=1}^M \sum_{\ell_3=1}^K G(\ell_1, \ell_2, \ell_3) u_{\ell_1 i} u_{\ell_2 j} u_{\ell_3 k}$$

で定義される。 $G \in \mathbb{R}^{N \times M \times K}$ はコアテンソル、 $u_{\ell_1 i} \in \mathbb{R}^{N \times N}$, $u_{\ell_2 j} \in \mathbb{R}^{M \times M}$, $u_{\ell_3 k} \in \mathbb{R}^{K \times K}$ は特異値行列である。特異値行列は直交行列である。まず、これは明らかに過完備であり、ユニークな答えは存在しない。したがって

タッカー分解の結果はタッカー分解を実行する具体的なアルゴリズムによって変わってしまう。ここでは高次特異値分解 [11] と呼ばれるアルゴリズムを採用する。高次特異値分解で得られたタッカー分解では一般にコアテンソルは対角テンソルではない。従って、特異値ベクトル $u_{\ell_1} \in \mathbb{R}^N$, $u_{\ell_2} \in \mathbb{R}^M$, $u_{\ell_3} \in \mathbb{R}^K$ は様々な組み合わせで複数回掛けあわされた上で和が取られる。

2.3 行列からのテンソル作成

本研究で実際にテンソル分解されるテンソルは、化合物を投与した際の遺伝子発現プロファイルを表現するテンソルまたはベクトルと、対象疾患の遺伝子発現プロファイルを表現するテンソルまたはベクトルから作られた合成テンソルになる。前者は DrugMatrix から取られるため $x_{j_1 j_2 i} \in \mathbb{R}^{D \times T \times N}$ という三相のテンソルであり、 i は遺伝子を、 j_1 は化合物を、 j_2 は化合物投与後の時刻を表す。後者はケースバイケースだがここでは簡単のため $x_{j_3 i} \in \mathbb{R}^{M \times N}$ というベクトル形式を取るものとしよう。ここで j_3 は患者と健常者を含むサンプルの ID であるとする。これらから四相のテンソル $\tilde{x}_{j_1 j_2 j_3 i} = x_{j_1 j_2 i} x_{j_3 i}$ を作ってこれをテンソル分解することにする。

$$\tilde{x}_{j_1 j_2 j_3 i} = \sum_{\ell_1=1}^D \sum_{\ell_2=1}^T \sum_{\ell_3=1}^M \sum_{\ell_4}^N G(\ell_1, \ell_2, \ell_3, \ell_4) u_{\ell_1 j_1} u_{\ell_2 j_2} u_{\ell_3 j_3} u_{\ell_4 i}$$

2.4 変数選択

テンソル分解の結果を使って変数選択する方法を説明する (図 1)。今回の解析の場合は、遺伝子発現プロファイルは、化合物投与後の時間依存性と健常者と患者の間の有意差を共に満たすようなものを選ぶ必要がある。これが疾患を対象とする化合物の同定につながる。そこで、化合物投与後の時間依存性を持つ特異値ベクトル $u_{\ell'_2 j_2}$ と、患者と健常者に差がある特異値ベクトル $u_{\ell'_3 j_3}$ が見つかったとしよう。これらと相補的な遺伝子特異値ベクトル $u_{i \ell_4}$ と化合物特異値ベクトル $u_{\ell_1 j_1}$ を見つけるために、 $G(\ell_1, \ell'_2, \ell'_3, \ell_4)$ を絶対値が大きい順に並べる。この順位が高い $u_{\ell'_4 i}$ と $u_{\ell'_1 j_1}$ が化合物投与後の時間依存性と健常者と患者の間の有意差に大きく寄与する化合物と遺伝子の組み合わせを探す鍵になる。

具体的には $u_{i \ell'_4}$ と $u_{\ell'_1 j_1}$ がガウス分布していることを仮定し、 χ 二乗分布を用いて

$$P_i = P_{\chi^2} \left[> \left(\frac{u_{\ell'_4 i}}{\sigma_{\ell'_4}} \right)^2 \right], P_{j_1} = P_{\chi^2} \left[> \left(\frac{u_{\ell'_1 j_1}}{\sigma_{\ell'_1}} \right)^2 \right]$$

の様に遺伝子 i 及び化合物 j_1 に与えられた P 値、 P_i 及び P_{j_1} を、Benjamin-Hochberg で多重補正して、補正 P 値が 0.01 以下である遺伝子と化合物を選択する。ここで $P_{\chi^2}[> x]$ は引数が x 以上である χ 二乗分布の累積確率分布、 $\sigma_{\ell'_4}, \sigma_{\ell'_1}$ は標準偏差である。

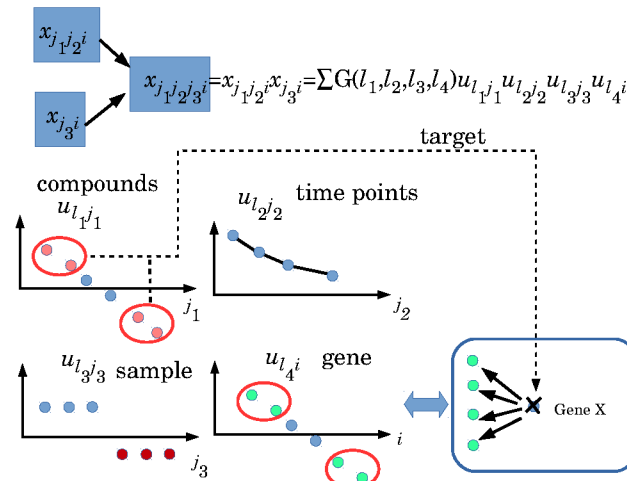


図 1 変数選択の模式図。DrugMatrix (GEO ID:GSE59927) から得られた、 j_1 番目の化合物がラットに投与されてから j_2 番目の時刻での i 番目の遺伝子の発現量を表すテンソルを $x_{j_1 j_2 i}$ とする。また、これとは独立に j_3 番目のサンプルの i 番目の遺伝子の発現プロファイルを表す行列 $x_{j_3 i}$ を用意する。このサンプルには典型的には健常者と患者のサンプルが含まれている。これらからテンソル $\hat{x}_{j_1 j_2 j_3 i}$ を $x_{j_1 j_2 i}$ と $x_{j_3 i}$ の単純な積で作る。次にテンソル $\hat{x}_{j_1 j_2 j_3 i}$ をテンソル分解して、化合物特異値ベクトル $u_{l_1 j_1}$ 、時刻特異値ベクトル $u_{l_2 j_2}$ 、サンプル特異値ベクトル $u_{l_3 j_3}$ 、遺伝子特異値ベクトル $u_{l_4 i}$ を得る。これらの中から以下の条件を満たす $\ell_k, 1 \leq k \leq 4$ を選ぶ。i) コアテンソル $G(\ell_1, \ell_2, \ell_3, \ell_4)$ が大きな絶対値をもつ。ii) 時刻特異値ベクトル $u_{l_2 j_2}$ が時間変化を伴っている、iii) サンプル特異値ベクトル $u_{l_3 j_3}$ が患者と健常者の差を表現している (朱色の円)。最後に、絶対値が十分大きなコアテンソル $G(\ell_1, \ell_2, \ell_3, \ell_4)$ が付随している、遺伝子特異値ベクトル $u_{l_4 i}$ と化合物特異値ベクトル $u_{l_1 j_1}$ を用いて、遺伝子 (薄緑色の円) と化合物 (ピンク色の円) を選択する。最後に、選択された遺伝子と、単一遺伝子をノックアウトした時の発現変化遺伝子のリストを照合し、一致するものを化合物の標的遺伝子として同定する。

Fig. 1 Intuitive illustration of the present strategy. Suppose there is a tensor, $\hat{x}_{j_1 j_2 j_3 i}$, which represents the i th gene expression at the j_2 th time point after the j_1 th compound is given to a rat; these data are taken from the DrugMatrix (GEO ID:GSE59927) dataset. There is also a matrix, $x_{j_3 i}$, which represents the i th gene expression of the j_3 th sample; samples typically include disease samples and control samples. Tensor $\hat{x}_{j_1 j_2 j_3 i}$ was generated as a “mathematical product” of $x_{j_1 j_2 i}$ and $x_{j_3 i}$. Then, tensor $\hat{x}_{j_1 j_2 j_3 i}$ is decomposed, and singular value matrix of compounds $u_{l_1 j_1}$, singular value matrix of time points $u_{l_2 j_2}$, sample singular value matrix $u_{l_3 j_3}$, and gene singular value matrix $u_{l_4 i}$, are obtained. Among them, I selected the combinations of $\ell_k, 1 \leq k \leq 4$, which are simultaneously associated with all of the following: i) core tensor $G(\ell_1, \ell_2, \ell_3, \ell_4)$ with a large enough absolute value, ii) a singular value vector of time points, $u_{l_2 j_2}$, whose value significantly varies with time, and iii) sample singular value vector $u_{l_3 j_3}$ associated with distinction between healthy controls and patients. These parameters are different between a disease (red filled circles) and control samples (cyan filled circles). Finally, using gene singular value vector $u_{l_4 i}$ and compound singular value vector $u_{l_1 j_1}$, compounds (filled pink circles) and genes (filled light-green circles) associated with $G(\ell_1, \ell_2, \ell_3, \ell_4)$ s with large enough absolute values are selected. Next, if the selected genes are coincident with the genes associated with a significant alteration when gene X is knocked out (or overexpressed), then the compounds are assumed to target gene X .

2.5 薬物標的タンパクの推定

前小節で選択した遺伝子は、あくまで化合物投与によって発現が変化する遺伝子であって、化合物の直接の標的で

はない。化合物の直接の標的はタンパクであり、遺伝子発現プロファイルは mRNA の量であるから、遺伝子発現プロファイルから直接化合物の標的タンパクを知ることはで

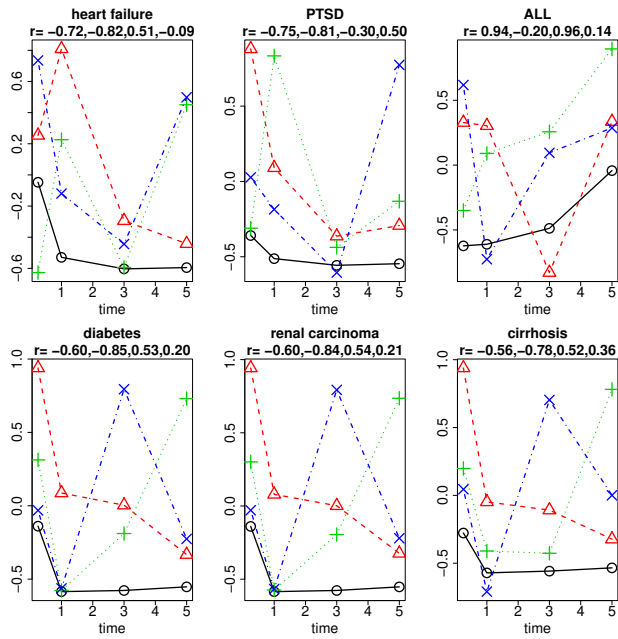


図 2 時刻特異値ベクトル。 r は観測時刻 (投与 1/4, 1, 3, 5 日後) と第一から第四時刻特異置ベクトル, $u_{\ell_2 j_2}$, $1 \leq j_2, \ell_2 \leq 4$, の間のピアソン相関係数。黒丸、赤三角、緑十字、及び、青バツ印は順に $\ell_2 = 1, 2, 3, 4$ である。

Fig. 2 Time point singular value vectors. r represents Pearson's correlation coefficients between time points (1/4, 1, 3, and 5 days after a treatment) and the first to fourth singular value vectors of time points, $u_{\ell_2 j_2}$, $1 \leq j_2, \ell_2 \leq 4$. Black open circles: $\ell_2 = 1$, red open triangles: $\ell_2 = 2$, green crosses: crosses: $\ell_2 = 4$. $j_2 = 1, 2, 3, 4$ correspond to time points 1/4, 1, 3, and 5 days, respectively.

きない。そこで、ここでは Enrichr [13] に実装されている、遺伝子を網羅的に KO した時の遺伝子発現プロファイルを参照することで、化合物の標的タンパクを推定する (図 1)。具体的には、前小節で選択した遺伝子を Enrichr にアップロードして、KO で影響を被る遺伝子との一致度が高い (補正 P 値で 0.01 以下) の遺伝子を化合物の直接の標的タンパクと推定することにする。

3. 結果

3.1 時刻特異値ベクトルの選択

図 2 に計算された時刻特異値ベクトルを示す。ALL 以外は $\ell_2 = 2$ が最も時刻と $u_{\ell_2 j_2}$ の間の相関係数の絶対値が大きいののでこれらを採用した。ALL だけは $\ell_2 = 3$ が最も時刻と $u_{\ell_2 j_2}$ の間の相関係数の絶対値が大きかったのでこれを採用した。

3.2 サンプル特異値ベクトルの選択

図 3 は、サンプル特異置ベクトルのカテゴリ / 時刻依存性である。いずれも有意な相関が出ている。これらに基づいて (A) 心臓疾患については $\ell_3 = 2, 3$ を、(B) PTSD については $\ell_3 = 1$ を、(C) ALL については $\ell_3 = 4$ を、

(D) 糖尿病については $\ell_3 = 1$ を、(E) 腎臓がんについては $\ell_3 = 13, 15, 30, 33, 35$ を、(F) 肝硬変については $\ell_3 = 2, 6$ をそれぞれ選択した。

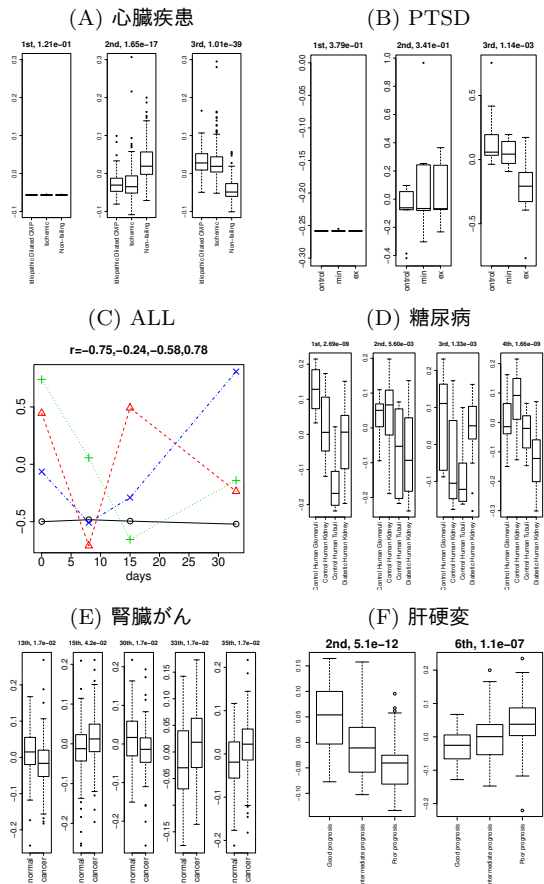


図 3 サンプル特異値ベクトルの箱ひげ図 ((A),(B),(D),(E),(F)) と時刻依存性 ((C))。箱ひげ図の上の P 値は (A),(B), および (D) については、カテゴリ回帰 (ANOVA と等価) の P 値、(E) と (F) については補正 P 値である。時刻依存性の図 ((C)) については相関係数を記載した。黒丸、赤三角、緑十字、及び、青バツ印は順に $\ell_3 = 1, 2, 3, 4$ である。

Fig. 3 Boxplots [for (A), (B), (D), (E), and (F)] and time dependence (C) for sample singular value vectors. The numbers above boxplots are P -values for (A), (B), and (D) and adjusted P -values for (E) and (F), computed by categorical regression (in other words, ANOVA). The numbers above time dependence (C) are correlation coefficients. (C) Black open circles: $\ell_3 = 1$, red open triangles: $\ell_3 = 2$, green crosses: $\ell_3 = 3$, and blue crosses: $\ell_3 = 4$. $j_3 = 1, 2, 3, 4$ correspond to time points 0, 8, 15, and 33 days, respectively.

3.3 遺伝子と化合物の選択、及び標的遺伝子の推定と評価

ここまでの結果に基づき、選択された ℓ_2, ℓ_3 を用いて、絶対値が相対的大きいという条件で $G(\ell_1, \ell_2, \ell_3, \ell_4)$ を選択し、この結果に基づいて遺伝子特異値ベクトル $u_{\ell_4 i}$ と化合物特異値ベクトル $u_{\ell_1 j_1}$ を選択し、その結果に基づい

て、遺伝子と化合物を選択した。さらに、選択された遺伝子から化合物が標的とする可能性の高い遺伝子を選択した(図 1, §2.5)。表 1 がその結果である(詳細は原論文 [1] 参照)。ALL を除くとすべて既知の化合物-標的タンパク関係と有意な相関を持った結果を出すことができていたことがわかる。

3.4 肝硬変治療薬の推定

この様な解析から実際に創薬候補化合物を推定できるかどうかを試すために、一般的には治療薬がないとされている肝硬変をターゲットに、この結果に基づいて薬剤候補化合物の推定を試みた。化合物では bezafibrate が、遺伝子では CYPOR と HNF4A がそれぞれ肝硬変に関する薬剤候補化合物とその標的遺伝子である可能性が高かった。bezafibrate が実際に CYPOR と HNF4A に結合するかどうかを SwissDock [14] で確認した。図 4 にその結果を示す。それぞれ Ki 値で 79nM と 0.13 μ M の値が出ており、阻害剤としての可能性は十分あるものと思われる。また、bezafibrate は過去に肝硬変の薬として検討されたことがあることも解った [15, 16]。実際に、bezafibrate が肝硬変の治療薬として認定されることは無かったのになんらかの問題点があったのだろうが、少なくとも提案手法が全く的外れではないという傍証にはなるだろう。

4. おわりに

本研究報告ではテンソル分解を用いた教師なし学習による変数選択を用いて、遺伝子発現プロファイルだけを用いて、既知の治療薬の情報を問わずに疾患の治療薬と標的遺伝子を推定する方法を提案した。その結果、検討した心臓疾患、PTSD, ALL, 糖尿病、腎臓がん、肝硬変の6つの疾患のうち、ALL を除く5つの疾患に対して、既知の治療薬-標的遺伝子関係と有意に相関する結果を出すことができた。また、特に治療薬がない肝硬変をとりあげ、この結果から治療薬が提案できるか試したところ、治療薬候補化合物である bezafibrate と候補標的タンパクである CYPOR と HNF4A には十分な結合能があることがドッキングシミュレーションで明らかになった。また、過去に bezafibrate は肝硬変の薬として検討されていたことも判明した。このようなことから提案手法は AI 創薬の一手段として有望であると結論づけた。

参考文献

- [1] Taguchi, Y.-H.: Identification of candidate drugs using tensor-decomposition-based unsupervised feature extraction in integrated analysis of gene expression between diseases and DrugMatrix datasets, *Scientific Reports*, Vol. 7, No. 1, p. 13733 (online), DOI: 10.1038/s41598-017-13003-0 (2017).
- [2] 執筆者: 79 名: *in silico* 創薬におけるスクリーニングの

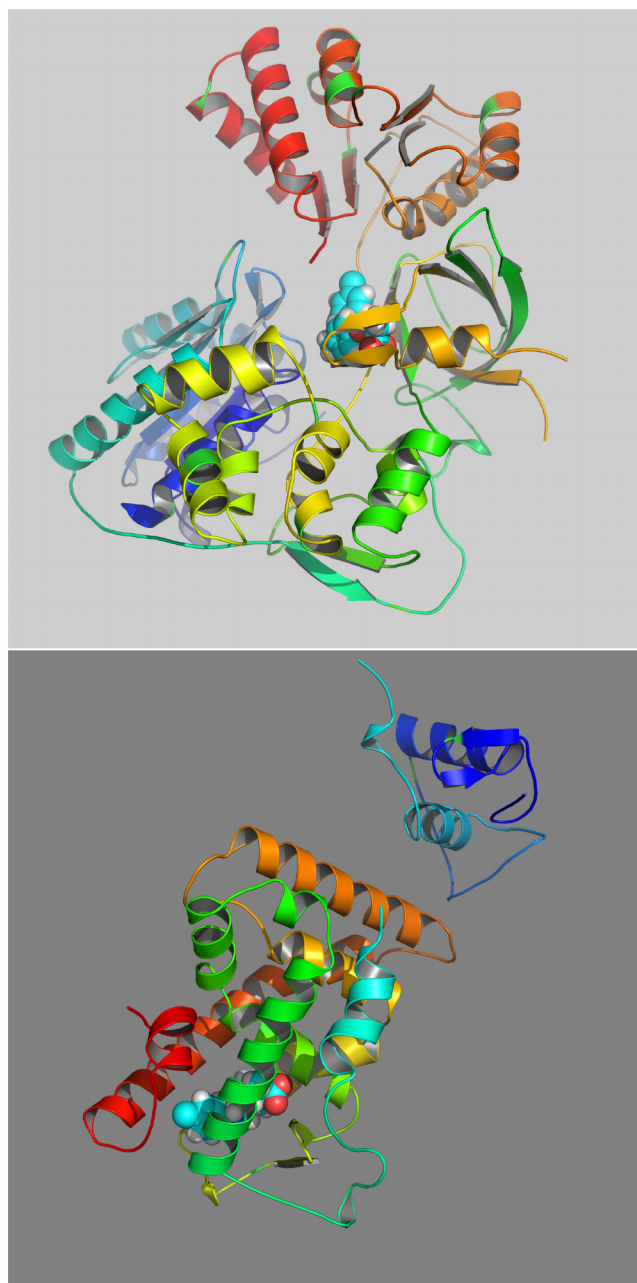


図 4 上: SwissDock で推定された bezafibrate と CYPOR の結合構造。下: SwissDock で推定された bezafibrate と HNF4A の結合構造。

Fig. 4 Upper: Binding structure of bezafibrate toward CYPOR computed by SwissDock. Lower: Binding structure of bezafibrate toward the ligand-binding domain of HNF4A computed by SwissDock.

- 高速化・高精度化技術, 技術情報協会, 初 edition (2018).
- [3] 山西 芳裕: オミックスデータから薬物の標的分子を網羅的に予測するための機械学習法の開発 (2016). <https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-25700029/>.
- [4] Taguchi, Y.-H. and Ng, K.-L.: Tensor decomposition-based unsupervised feature extraction for integrated analysis of TCGA data on microRNA expression and promoter methylation of genes in ovarian cancer, (online), DOI: 10.1101/380071 (2018).
- [5] Taguchi, Y.-H.: Tensor Decomposition-Based Unsuper-

表 1 化合物の既知標的タンパクと、本研究の予測標的タンパクとの間のフィッシャーの正確確率検定 (P_F) と χ^2 検定 (P_{χ^2}) の結果。行: DINIES [12] による化合物既知標的。列: Enrichr [13] の “Single Gene Perturbations from GEO up” または “Single Gene Perturbations from GEO down” による標的タンパク推定結果。OR はオッズ比。

Table 1 Fisher’s exact test (P_F) and the uncorrected χ^2 test (P_{χ^2}) of known drug target proteins regarding the inference of the present study. Rows: known drug target proteins (DINIES [12]). Columns: Inferred drug target proteins using ‘Single Gene Perturbations from GEO up’ or ‘Single Gene Perturbations from GEO down’. OR: odds ratio

		Single Gene Perturbations from GEO up					Single Gene Perturbations from GEO down				
		F	T	P_F	P_{χ^2}	RO	F	T	P_F	P_{χ^2}	RO
heart failure	F	521	517	3.4×10^{-4}	3.9×10^{-4}	3.02	628	416	1.3×10^{-3}	7.3×10^{-4}	2.61
	T	13	39				19	33			
PTSD	F	500	560	3.8×10^{-2}	3.1×10^{-2}	2.67	532	529	6.1×10^{-3}	4.5×10^{-3}	3.81
	T	6	18				5	19			
ALL	F	979	89	2.7×10^{-1}	3.0×10^{-1}	2.19	1009	57	1.0×10^0	-	-
	T	10	2				12	0			
diabetes	F	889	177	1.2×10^{-2}	7.1×10^{-3}	3.00	936	130	3.6×10^{-4}	2.0×10^{-5}	5.13
	T	15	9				14	10			
renal carcinoma	F	847	219	2.0×10^{-2}	1.2×10^{-2}	2.75	895	169	4.3×10^{-2}	2.2×10^{-2}	2.64
	T	14	10				16	8			
cirrhosis	F	572	219	1.1×10^{-2}	8.1×10^{-3}	2.91	595	169	1.6×10^{-3}	1.1×10^{-3}	3.81
	T	8	10				7	8			

vised Feature Extraction Can Identify the Universal Nature of Sequence-Nonspecific Off-Target Regulation of mRNA Mediated by MicroRNA Transfection, *Cells*, Vol. 7, No. 6, p. 54 (online), DOI: 10.3390/cells7060054 (2018).

- [6] Taguchi, Y.-H.: Tensor decomposition-based and principal-component-analysis-based unsupervised feature extraction applied to the gene expression and methylation profiles in the brains of social insects with multiple castes, *BMC Bioinformatics*, Vol. 19, No. S4 (online), DOI: 10.1186/s12859-018-2068-7 (2018).
- [7] Taguchi, Y.-H.: Tensor decomposition-based unsupervised feature extraction identifies candidate genes that induce post-traumatic stress disorder-mediated heart diseases, *BMC Medical Genomics*, Vol. 10, No. S4 (online), DOI: 10.1186/s12920-017-0302-1 (2017).
- [8] Taguchi, Y.-H.: Identification of Candidate Drugs for Heart Failure Using Tensor Decomposition-Based Unsupervised Feature Extraction Applied to Integrated Analysis of Gene Expression Between Heart Failure and Drug-Matrix Datasets, *Intelligent Computing Theories and Application*, Springer International Publishing, pp. 517–528 (online), DOI: 10.1007/978-3-319-63312-1_45 (2017).
- [9] Taguchi, Y.-H.: Tensor decomposition-based unsupervised feature extraction applied to matrix products for multi-view data processing, *PLOS ONE*, Vol. 12, No. 8, p. e0183933 (online), DOI: 10.1371/journal.pone.0183933 (2017).
- [10] Taguchi, Y.-H.: One-class Differential Expression Analysis using Tensor Decomposition-based Unsupervised Feature Extraction Applied to Integrated Analysis of Multiple Omics Data from 26 Lung Adenocarcinoma Cell Lines, *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, IEEE, (online), DOI: 10.1109/bibe.2017.00-66 (2017).
- [11] 石黒勝彦, 林 浩平: 関係データ学習 (機械学習プロフェッ

ショナルシリーズ), 講談社 (2016).

- [12] Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M. and Goto, S.: DINIES: drug-target interaction network inference engine based on supervised analysis, *Nucleic Acids Res.*, Vol. 42, No. Web Server issue, pp. 39–45 (2014).
- [13] Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W. and Ma’ayan, A.: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Res.*, Vol. 44, No. W1, pp. W90–97 (2016).
- [14] Grosdidier, A., Zoete, V. and Michielin, O.: SwissDock, a protein-small molecule docking web service based on EADock DSS, *Nucleic Acids Res.*, Vol. 39, No. Web Server issue, pp. W270–277 (2011).
- [15] 金子 晃, 久保光彦, 渡辺晋一郎, 東谷光庸, 山本守敏, 巽 信之, 中間昭弘, 尾下正秀, 片山和宏, 井上敦雄, 春名能通, 吉原治正, 久保田真司, 三田英治, 鈴木都男, 加藤道夫, 脇岡泰三, 萩原秀紀, 平松直樹, 林 紀夫: 原発性胆汁性肝硬変に対する bezafibrate 療法の有効性と問題点, *肝臓*, Vol. 46, No. 4, pp. 200–207 (オンライン), DOI: 10.2957/kanzo.46.200 (2005).
- [16] 矢野公士, 加藤浩之, 森田茂樹, 古河隆二: 原発性胆汁性肝硬変に対する bezafibrate の長期投与効果, *肝臓*, Vol. 43, No. 2, pp. 115–120 (オンライン), DOI: 10.2957/kanzo.43.115 (2002).