

博士論文（要約）

非負値行列因子分解を用いた  
薬剤の効果予測バイオマーカーの探索

藤田 直也

## 背景と目的

分子標的薬と効果予測バイオマーカーを用いたがん患者に対する精密医療(Precision medicine)は従来型の全患者共通の医療から各患者へ適した医療へとパラダイムシフトをもたらした。バイオマーカーを用いて適切な患者を選び適切な治療法を提供することは、副作用のリスクを下げ薬効が見込める患者のみを選ぶことを可能にし、薬剤の価値を高めると考えられる。したがって、そのような薬効を予測するバイオマーカーを同定することが強く求められている。

バイオマーカー探索を目的として、がん細胞株を用いて薬効と関連するゲノム異常を同定する試みがなされてきた。数百細胞株にも及ぶ大規模な細胞株パネルを対象に、様々な薬剤への感受性ならびに、変異や発現量等のマルチオミクス情報を取得し、ANOVA や  $t$  検定等の統計的手法、あるいは Elastic net 回帰分析等の回帰モデルを用いてバイオマーカーと薬剤標的分子の関係を明らかにできることが報告されている。しかし、前者の手法は個々のペアの評価に留まり、後者も各薬剤に対する予測モデルであるため、薬剤間の特徴比較等、全データを対象とした包括的な解釈には限界があった。またマルチカーネル法等の高度な機械学習手法を用いて薬効を予測する試みも進んでいるが、マルチオミクス全てを入力とした予測モデルであるため、臨床応用に際し全てを測定することのハードルは高いと思われる。

包括的な解釈にはクラスタリングが従来からの有効なアプローチである。その中でも、非負値行列因子分解(Non-negative matrix factorization, NMF)は明示的なクラスタリングではなく潜在的な要素に分解することで行列要素間の関連性を際立たせることができ、さらには、変異や過剰発現等の外れ値を捉えることに適した手法である。また NMF を拡張した結合非負値行列因子分解(Joint NMF, JNMF)が提案されている[1]。JNMF は複数行列を対象にすることができマルチオミクスデータの統合解析を可能とする。

本研究では、薬理学データを含むマルチオミクス情報に対して JNMF を実行し薬剤感受性とゲノム異常の関係性を明らかにすることを試みた。続いて、パスウェイ解析を実施し単なるクラスターの相関関係に解釈を加え、さらには、遺伝子シグニチャー解析でパスウェイ活性度指標を作成することにより、薬効とパスウェイ活性度の関連性を考察した。

## データと方法

マルチオミクスデータとして Cancer Cell Line Encyclopedia (CCLE) から化合物感受性、変異、コピー数増幅と欠失、mRNA 発現量、がん種の 6 種類の情報を取得した。感受性情報を含む細胞株数は 504 となった。また、方法の妥当性検討用のデータとして人工的に 3 つの行列を作成し、化合物感受性、mRNA 発現量、変異のデータと見立てて用いた。解析手法は JNMF を採用し、 $N$  個の入力行列  $X_I$  を  $WH_I$  で再現する部分行列  $W$  と  $H_I$  を求める。

$$\min \sum_{I=1}^N \|X_I - WH_I\|_F^2$$

ここで、 $W$ は全てで共通する行列で、 $H_i$ は個々の入力行列 $X_i$ に対応し生成される行列である。さらに今回、欠測値を含むデータに対応できるようにマスク行列を用いて欠測の行列要素を評価しないように改良した。この JNMF 解析法を用いて、オミクスデータ間で共通のクラスターに所属する特徴量（コモジュール）の抽出を実施した。結果の再現性はコンセンサス行列ならびに複数の試行の一致度を示すコーフェン相関係数で確認した。パスウェイ解析は Ingenuity Pathway Analysis (IPA) [2]を用いた。また遺伝子発現量シグニチャーの尺度として平均発現量を用いた。

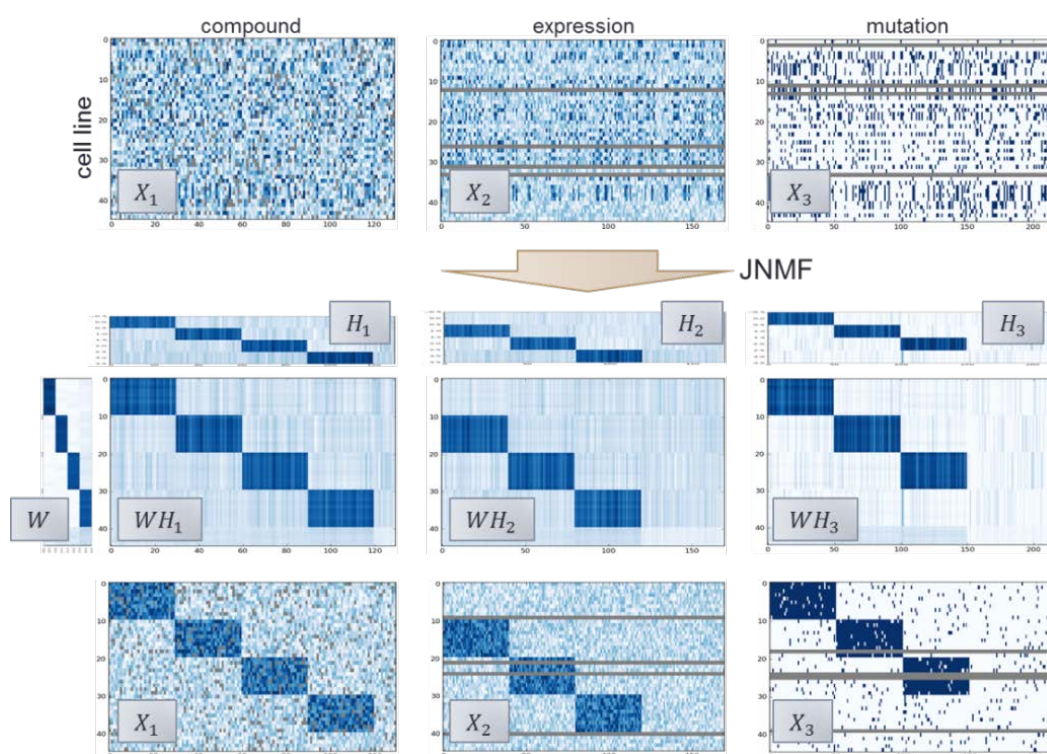


図1. 人工的な複数の入力行列に対して JNMF を実行した結果

## 結果と考察

最初に、人工的なデータを用いて JNMF を実施し、期待通りのコモジュールが検出できるかを確認した。図1に示す通り、3つの入力 $X_1$ 、 $X_2$ 、 $X_3$ をよく再現する $W$ と $H_1$ 、 $H_2$ 、 $H_3$ を同定することができた。 $W$ と $H$ のランクを4とあらかじめ定義し JNMF を実行することで、4つのコモジュールを選び出すことができている。つまり JNMF を用いることで、より低次元の部分行列である $W$ と $H$ で入力 $X$ を表現できていることを示している。加えて、生物学データでよく直面する欠測値にも対処できるよう JNMF アルゴリズムを改良することで、今回の人工的なデータにおいては欠測値が含まれている場合でも正しくコモジュールが検出できることを見いだした。同時に、欠測の値を正しく推定することも明らかにした。これは、複数のオミクスデータを組み合わせることで未計測の遺伝子変異や薬剤感受性を予測できる可能性を示す。

次に、実際の大規模細胞株パネル CCLE から得られた薬剤感受性プロファイルと網羅的オミクスデータ、がん種からなる多次元的情報を入力として JNMF を実行した。様々なランク  $k$  で実施し、解が局所的に安定であった 40 の場合に着目した。10 回の JNMF 試行から作成した  $W$  に対するコンセンサス行列のコーフェン相関係数は 0.91 であり、このことから、細胞株群を概ね再現性高く 40 のクラスターに分類できることを確認した。さらに、それらのクラスターに対応するオミクス特徴量のプロファイル  $H$  を解析し、既知のバイオマーカーと薬剤の標的分子の関係性を抽出できることを確認した。例えば EGFR 阻害感受性と EGFR 過剰発現が同じコモジュールに所属することを確認した。加えて、EGFR 阻害と HER2 阻害のプロファイルの違いとして、HER2 阻害感受性のみ HER2 遺伝子増幅ともコモジュールができることを確認した。つまり JNMF 解析により、類似の ErbB ファミリーを阻害する EGFR 阻害剤と HER2 阻害剤の違いを際立たせることができ、かつそれらの違いをバイオマーカーで解釈できることを見いだした。薬剤の差別化戦略の立案に有効と考える。

さらに、新規のバイオマーカーとして MITF 活性度の可能性を見いだした。JNMF クラスターの中でも特に再現性が高かったメラノーマのクラスターに着目した。それには、BRAF 阻害剤 PLX4720、BRAF 変異、メラノーマがコモジュールとして所属している。このコモジュールの解釈を深めるために、過剰発現遺伝子群を対象として IPA パスウェイ解析を実施し、共通に制御する発現制御因子の同定を試みた。結果、数多くの過剰発現遺伝子が転写因子 MITF に共通に制御されていることを見いだした (IPA overlap,  $p = 6.91e-39$ )。MITF 遺伝子増幅も所属することも含めて、このコモジュールは MITF の活性化を表現していると示唆された。BRAF の下流に MITF があるという今までの知見とも合致する。続いて、その MITF 活性化と BRAF 阻害感受性に関連性があるかを確認した、MITF 制御下遺伝子の平均発現量を MITF 活性化シグニチャーと見なし、BRAF 阻害感受性と BRAF 変異、予測された MITF 活性度の関連性を調べた。その結果、PLX4720 感受性は BRAF 変異かつ MITF 活性化の群に偏っており (Kruskal-Wallis test,  $p = 3.44e-05$ )、既知バイオマーカーである BRAF 遺伝子に変異があるが MITF が活性化していないメラノーマ細胞株群には感受性ではないものも含まれていた (図 2)。したがって、MITF 活性化シグニチャーと BRAF 変異を組み合わせると BRAF 阻害剤が効くがんをより絞って集められる可能性が示唆された。

## 結論

今回、JNMF とパスウェイ解析を組み合わせることで細胞株のマルチオミクスデータを解析することにより、MITF 活性度が新たなバイオマーカーになる可能性を見いだした。BRAF 阻害剤に対して BRAF 変異は効果予測バイオマーカーとして確立されているが、MITF の活性度を加味する方がより感受性株を選び出すことができた。BRAF/MITF パスウェイが上流から下流まで活性化しているがんに対して、BRAF 阻害剤が有効であると解釈できる。

また、バイオマーカー提案と同時に薬剤の差別化戦略を立案できる可能性も見いだされた。ErbB ファミリー阻害剤である EGFR 阻害剤と HER2 阻害剤のプロファイルの違いを JNMF の結果で示すことができた。数万のマルチオミクス特徴量の中から、両薬剤間の共通性と相違性が EGFR 過剰発現、HER2 増幅といった既知のバイオマーカーと同じコモジュールとして選び出されたことが重要である。

上記2つの知見を示したことにより、本アプローチはマルチオミクスの複数の入力情報を取り扱い(multi-modal)、バイオマーカー発見と薬剤差別化戦略の複数の目的(multi-task)を同時に達成することができる方法論と見なすことができる。2つの部分行列に分解するという NMF の単純な性質は、スパースなデータであり解釈も求められる医学の分野で有効であると思われる。また、NMF は単純な手法であるため、細胞株に限定せず、クリニカルシーケンスや血中 cell free DNA を用いたマーカーモニタリング等、臨床の網羅的オミクスデータにも応用できる可能性がある。さらに、一研究機関で閉じたデータだけに限らず、国際がんゲノムプロジェクトのようなリファレンスと比較して判断する時代になり、大規模データを統合的に解釈する必要性は高い。そのようなマルチオミクス解析の方法として JNMF は有効であると考えられる。

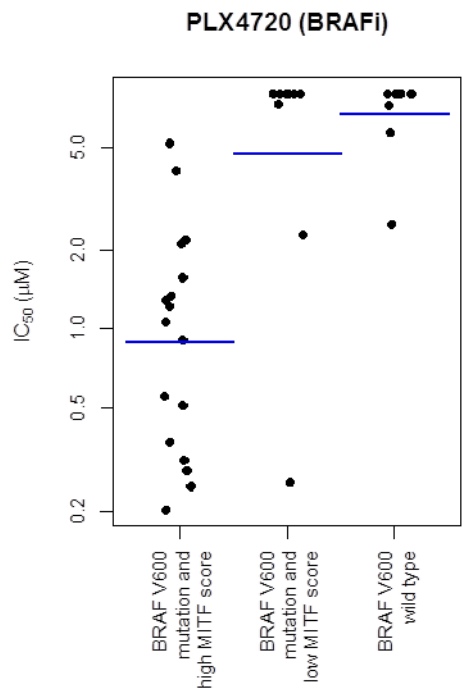


図2. MITF 予測活性度と BRAF 変異で分類したメラノーマ細胞株群と BRAF 阻害剤感受性の関係

[1] Zhang, *et al.*, *Nucleic Acids Res.* 2012 Oct;40(19):9379-91.  
 [2] Krämer, *et al.*, *Bioinformatics.* 2014 Feb 15;30(4):523-30.