G-008

# 遺伝子発現情報の機械学習に基づく投薬効果予測

# Prediction of drug therapy efficacy by support vector machine based on the personal gene expression patterns

吉田　寛輝† 峯田　克彦† 遠藤　俊徳†
Hiroki Yoshida Katsuhiko Mineta Toshinori Endo

## 1. 背景と目的

*in vitro* での薬剤有効率と *in vivo* での薬剤有効率とでは、その環境の違いにより大きく違い、*in vitro* では有効率が高いものでも *in vivo* では有効率が低くなる薬剤は多い。また、使用薬剤の決定も経験的判断による部分が大きく、時として無駄な投薬や重篤な副作用をもたらす可能性がある[1]。投薬前に対象患者に対してその薬剤の効果を予測し、適切な薬剤を選択することが出来れば、患者の生活の質の向上に繋がる。また、マイクロアレイによる臨床検体を対象とした解析は多くの解析方法が示されているが、未だに暗中模索の感がある[2]。本研究では、マイクロアレイを利用した遺伝子発現データからの遺伝子の絞り込み選択方法と、最適な学習判別器の構成方法を研究することにより、実際に臨床に応用できる投薬効果の予測判別方法の確立、及び、自己免疫疾患における薬効メカニズムを解明することを目的とした。また、本研究においては、精度が高いことで注目されているサポートベクターマシン(SVM)を学習判別機械とする投薬効果予測方法の確立を行った。

## 2. 材料と方法

匿名化した自己免疫疾患患者における、末梢血検体についてのマイクロアレイを用いた遺伝子発現量、及び、医師による各患者に対する投薬効果判定結果を用いた。患者数は 73 人であり、遺伝子数は 789 個である。そのうちの 66 人は学習データとして、7 人は判別データとして用いた。投薬効果判定は、アメリカリウマチ学会が定めている投薬効果判定基準の ACR[3]を基に判定されている。

学習データに用いた患者 66 人のうち、ACR70 もしくは ACR50 と判定された患者 32 人を投薬効果ありの群として、ACR20 もしくは ACR0 と判定された患者 34 人を投薬効果なしの群として、2 群に分けた。

これらのデータをもとに *t* 検定を用いて、遺伝子の絞り込み選択を行った。多重検定の場合、偽陽性である遺伝子が採用されてしまう恐れがあるため、偽陽性と考えられうる最大遺伝子数を *p*-value の低い順に除いた。

選択された遺伝子について、統計ソフト R(ver.2.7.2)で SVM のパラメータ最適化を行い、学習データを用いて学習判別の精度とサポートベクトル数を調べた。最適化は R の tune 関数を用い、精度評価は leave one out 法を用いた。用いたカーネルは Gaussian, Linear, Polynomial, Sigmoid の 4 つであり、用いたタイプは C-classification, Nu-classification の 2 つである。用いたカーネルの式は以下の通りである。

Gaussian: $d(x,y) = exp(-gamma^*|x-y|^2)$

Linear : マッピングは行わず、元の特徴空間で線形分離

Polynomial : $d(x,y) = (gamma^*(x \cdot y)+coef0)^{degree}$

Sigmoid : $d(x,y) = tanh(gamma^*(x \cdot y)+coef0)$

続いて、学習データには用いられていない 7 人（投薬効果ありの群は 5 人、投薬効果なしの群は 2 人）の患者を判別データとして用い、学習データを用いて構築したモデルの、未知データに対する判別能力（汎化能力）を調べた。

次に、偽陽性を考慮して選択された遺伝子の機能を調べ、自己免疫疾患における投薬効果の有無を決定づける遺伝子の特徴を明らかにした。

最後に、薬効メカニズムを調べるために、投薬効果の見られた患者 32 人の投薬前後の遺伝子発現量の変化を調べた。まず、*t* 検定を用いて投薬前後で有意に変化した遺伝子を調べた。今回も多重検定であるので、FDR をコントロールする方法[4]により、偽陽性を考慮した。この方法は、新たな閾値 $\alpha$ を求めることで、*p*-value$<\alpha$ となった時、統計的に有意とする方法である。$\alpha$ を求める式は以下の通りである。

$\alpha = (0.05 \times i)/$全遺伝子数　($i$: $p$ 値が低い方からの順位)

## 3. 結果と考察

### 3.1 *t* 検定の結果

*t* 検定の結果、*p*-value<0.05 で選択された遺伝子は 49 個であった。全部で 789 個の遺伝子を調べたので、今回、偽陽性と考えられる遺伝子の最大数は、$789 \times 0.05 = 39.45$ となり、*p*-value の低い順に 40 個の遺伝子を除いた。その結果、残った遺伝子は以下の 9 遺伝子である（表 1）。

**表 1　選択された遺伝子**

| 遺伝子名 | 機能 |
|---|---|
| Asialoglycoprotein receptor 1* | 損傷したグリコタンパクを除去(エンドサイトーシス) |
| Nonmetastatic protein 23* | 造血幹細胞で、分化を抑制しアポトーシスを促進 |
| Transforming growth factor, beta 1* | 組織修復 、炎症反応や免疫反応を終了させる調節因子 |
| DNA fragmentation factor 40K, beta* | アポトーシスの最期の段階における、DNA の分解 |
| Pyridoxal (pyridoxine, vitamin B6) kinase | ビタミン B6 が PLP になる際のリン酸化を促進 |
| CX3C chemokine receptor 1* | 血管内から炎症組織への白血球の遊走 |
| t-complex-associated-testis-expressed 1-like 1 | 人の dynein のサブユニット |
| DNA fragmentation factor 45K, beta* | アポトーシスの最期の段階における、DNA の分解 |
| ERK activator kinase | ERK を活性化させる |

† 北海道大学大学院 情報科学研究科 Graduate School of Information Science and Technology, Hokkaido University

### 3.2 SVM の結果

選択された 9 遺伝子を用い、1 患者ごとに 9 次元データのベクトルとして、全 66 ベクトルが正しく判別されたかを leave one out 法を用いて精度評価した。SVM の学習判別の結果、C-classification タイプと Nu-classification タイプにおける、Linear カーネルと Gaussian カーネルのそれぞれの組み合わせ（計 4 種類の組み合わせ）において、100%の精度を示した。また、C-classification タイプと Linear カーネルの組み合わせによって最少のサポートベクトル数 21 を示した。計算量の観点から、サポートベクトル数は少ないほどよいモデルとされており、また、一般的な SVM モデルのサポートベクトル数は全ベクトル数の 10%～30%である[5][6]。この結果より、C-classification タイプと Linear カーネルの組み合わせのモデルのサポートベクトル数は、学習データの全 66 ベクトルの約 30%であり、一般的な SVM モデルの範囲内にあることが分かった。

次に、学習データを用いて構築したモデルを用いて、未知データの判別能力（汎化能力）を調べた。結果は、7 人中何人が正しく判別されたかを示している（表 2）。

表 2　汎化能力結果

| | C-classification | Nu-classification |
|---|---|---|
| Gaussian | 100% | 100% |
| Linear | 100% | 100% |
| Polynomial | 42.9% | 85.7% |
| sigmoid | 71.4% | N.D. |

表 2 より、C-classification タイプと Nu-classification タイプにおける、Linear カーネルと Gaussian カーネルのそれぞれの組み合わせにおいて、100%の汎化能力を示した。よって、自己免疫疾患における投薬効果の予測に必要十分な 9 遺伝子を同定できた。

### 3.3 9 遺伝子の特徴

表 1 を見ると、9 遺伝子中 6 遺伝子がアポトーシスやエンドサイトーシスや組織修復、免疫に関する遺伝子であることが分かる(*印のついた遺伝子)。よって、自己免疫疾患における投薬効果の有無を決定づける遺伝子の特徴は、アポトーシスやエンドサイトーシスによって不要な細胞を除いて事前に炎症を防ぐか、炎症した組織を修復する遺伝子が働いているかで決定していると示唆される。

### 3.4 薬効メカニズム

*t* 検定のみ用いた時、*p*-value<0.05 で有意に変化した遺伝子は 127 個であり、投薬後に発現量が有意に低下したのは 127 遺伝子中わずかに 3 遺伝子であった。また、FDR をコントロールする方法を用いて遺伝子選択を行った結果、22 遺伝子が有意に変化した。全 22 遺伝子において投薬後に発現量が上昇した。22 遺伝子のうち、14 遺伝子は ribosomal protein であった。残りの 8 遺伝子は表 3 の通りである。

表 3　ribosomal protein 以外の遺伝子

| 遺伝子名 |
|---|
| T-cell cyclophilin |
| HLA class II DP beta 1 |
| HLA class II DQ beta |
| immunoglobulin kappa |
| stathmin |
| HLA class II DM alpha |
| Liver-specific cytochrome c oxidase (COX VIIa-L) |
| Immunoglobrin rearranged gamma chain |

このことから、ribosomal protein の発現量上昇により、これらのタンパク合成を促進していると考えられる。

表 3 の遺伝子から、Ig や HLA class II など、免疫に関わる遺伝子が多くみられる。自己免疫疾患は自己抗体の過量産生や、自己組織に反応する T 細胞発現が原因であり、投薬後に効果が見られた患者の免疫遺伝子の発現が上昇したことは興味深い。これらから、投与された低分子化合物がハプテンとして働き、Ig 産生を促し、それによって間接的に T 細胞活性化を誘導したと考えられる。ここで活性化された T 細胞は、免疫反応の対象を自己組織から別の対象に移行しているのかもしれない[7]。

また、COX VIIa-L 遺伝子は、年齢を重ねるにつれてプロモータ領域のメチル化が進み、遺伝子発現量が低下することで、2 型糖尿病に関与していることが報告されている[8]。2 型糖尿病と自己免疫疾患は、年齢を重ねるにつれて発病しやすいという表面的類似性がある。このことより、COX VIIa-L 遺伝子の抑制は自己免疫疾患にも関係しているかもしれない。

## 4. 結論

本研究により、自己免疫疾患における投薬効果の有無を判別するのに必要十分な 9 遺伝子を同定することが出来た（表 1）。

SVM のパラメータ最適化を行い、C-classification タイプと linear カーネルの組み合わせを使い、表 1 に示す 9 遺伝子を用いることで、患者負担の少ない、わずかな末梢血液細胞の遺伝子発現量検出による効果的かつ簡便な投薬効果予測を行うことが出来た。

自己免疫疾患における薬効メカニズムは、薬がハプテンとなって Ig 産生を促し、間接的に T 細胞活性化を促進することであると考えられる。ribosomal protein と免疫関連分子の発現上昇は、このことを示唆している。

また、2 型糖尿病に関係している COX VIIa-L 遺伝子が、自己免疫疾患に関わっている可能性があることが分かった。

### 参考文献

[1]長谷川清ら；マイクロアレイデータの多変量解析による抗癌剤感受性規定因子の選択および未知検体の感受性予測；中外製薬株式会社, 2007.
[2]佐藤陽美；分子ネットワークとチップデータ解析の融合～KeyMolent の開発～；バイオテクノロジージャーナル,2005.
[3]Frank C. Arnett *et al*; The american rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis, 1988.
[4]Koen J.F. Verhoeven *et al*: Implementing false discovery rate control: increasing your power, 2005.
[5]Davide AnguitaRidella *et al*; An Algorithm for Reducing the Number of SupportVectors, 2005.
[6]Quang-AnhTrainZhang *et al*; Reduce the Number of Support Vectors byUsing Clustering Techniques, 2003.
[7]Shunichi Kumagai *et al*; Immune Responses to Hapten-Modified Self and Their Regulation in Normal Individuals and Patients with SLE, 1981.
[8]Rönn T *et al*; Age influence DNA methylation and gene expression of COX7A1 in human skeletal muscle,2008.