

博士学位論文審査要旨

2022年1月20日

論文題目： Prediction method for therapeutic response at multiple time points of gene expression profiles

(時系列の遺伝子発現プロファイルを使用した治療効果の予測方法)

学位申請者： 福島 亜梨花

審査委員：

主査： 生命医科学研究科 教授 廣安 知之

副査： 生命医科学研究科 准教授 日和 悟

副査： 東京医科大学医学部 教授 杉本 昌弘

要 旨：

バイオマーカーはある特定の疾病や体質に関与する生体内分子のことであり、適切なバイオマーカーが発見されれば、患者個人の体質に合った治療を提供する個別化医療の実現や創薬分野での研究開発効率の向上が期待される。バイオマーカーを探索するためには、生体分子の網羅的な解析が必要となる。DNA (deoxyribonucleic acid) マイクロアレイは、1回の実験で数千から数万の遺伝子発現量を包括的に計測することができるが、遺伝子発現量はダイナミクスを持つため、複数の時点で計測した時系列のマイクロアレイデータは、単一の時点のものと比較して、よりバイオマーカーの探索に貢献できると考えられる。これを実現するための課題は、(1) 収集した数千から数万の遺伝子発現量に対してサンプルとなる患者の数が圧倒的に少ないことにより探索が困難になること、(2) 遺伝子同士の発現量が高い相関関係にある場合に誤ったバイオマーカー探索につながる可能性があること、(3) 遺伝子発現量は常に変化しているために計測するタイミングによっては精度が低くなることの3つの課題がある。本研究では、時系列のマイクロアレイデータから遺伝子発現量を用いたバイオマーカーを探索する、expanding elastic net for consistent differentiation (eENCD) と consolidating probabilities of multiple time points (CPMTP) を開発し、その有効性を示している。本論文は5章で構成され、各結果についてまとめている。

1章では、本研究の背景と動機について述べている。

2章では、治療効果の有無によって発現量が一貫して異なる値を示す遺伝子を選択できるeENCDの詳細や評価に関して述べており、バイオマーカー探索手法としてのeENCDの有効性を示した。

3章では、治療効果の有無において発現パターンが異なる遺伝子を選択する手法であるCPMTPの詳細や評価に関して述べており、バイオマーカー探索手法としてのCPMTPの有効性を示した。

4章では、eENCDとCPMTPの比較について述べ、CPMTPはeENCDよりも有効である可能性を示唆した。

5章では、本研究で得られた結論を述べた。

本論文の研究成果により、eENCDとCPMTPは、治療効果を正確に予測し、時系列のマイクロアレイデータの膨大な遺伝子の中から、標的の疾患に関連する遺伝子を選択することができることを示した。従って、本研究はバイオマーカーを用いた個別化医療の確立に貢献し、バイオイ

ンフォマティクス分野において多大に貢献することが期待できる。よって、本論文は、博士（工学）（同志社大学）の学位を授与するにふさわしいものであると認められる。

総合試験結果の要旨

2022 年 1 月 20 日

論文題目： Prediction method for therapeutic response at multiple time points of gene expression profiles

(時系列の遺伝子発現プロファイルを使用した治療効果の予測方法)

学位申請者： 福島 亜梨花

審査委員：

主 査： 生命医科学研究科 教授 廣安 知之

副 査： 生命医科学研究科 准教授 日和 悟

副 査： 東京医科大学医学部 教授 杉本 昌弘

要 旨：

本論文提出者は、2016 年 4 月に同志社大学大学院生命医科学研究科医工学・医情報学専攻博士課程後期課程に入学し、優れた研究成果を挙げている。また、本研究科修了に必要な所定の単位を取得するとともに、英語の語学試験にも合格し、その能力についても十分であると認定されている。

本論文の主要部分は、Scientific Reports 誌（2019 年）、BMC Bioinformatics 誌（2021 年）に掲載され、高い評価を受けている。さらに、その他の業績として、4 件の査読論文（内 3 件は共著者）、5 件の国際会議発表、7 件の口頭発表を行っている。

2022 年 1 月 20 日午後 1 時 30 分から 1 時間半にわたり提出論文に関する学術講演会（博士論文公聴会）が開催され、提出者の講演と種々の質疑応答がなされたが、提出者の説明により研究内容が学問的に優れていることが確認できた。公聴会終了後、提出論文に関係した学問的諸問題について、審査委員による口頭試問を実施し、提出者が博士学位授与にふさわしい能力を有することを確認した。よって、総合試験の結果は合格であると認める。

博士学位論文要旨

論文題目： Prediction method for therapeutic response at multiple time points of gene expression profiles

(時系列の遺伝子発現プロファイルを使用した治療効果の予測方法)

氏 名： 福島 亜梨花

要 旨：

バイオマーカーはある特定の疾病や体質に関与する生体内分子のことであり、薬力学、予後、予測マーカーの3種類に分類される。薬力学バイオマーカーは、新しい治療薬の開発に使用されており、Histone γ H2AX などが例として挙げられる。予後バイオマーカーは、疾患の再発や進行、死亡など予後に関する情報を提供する。例えば、Oncotype DX™は21個の遺伝子の発現量によって乳がんの再発スコアを計算する。予測バイオマーカーは、治療の効果や副作用の有無を治療前に提供できる。最も有名な予測バイオマーカーの1つはPD-L1であり、トリプルネガティブ乳がんの患者のPD-L1に関連する遺伝子の発現量を測定することで、免疫チェックポイント阻害剤による治療効果を予測することができる。このように、バイオマーカーとして遺伝子発現量を測定することは、患者の治療戦略を決めるために有益な情報をもたらしている。適切なバイオマーカーが発見されれば、患者個人の体質に合った治療を提供する個別化医療が実現でき、医療分野における質の向上が見込め、創薬分野での研究開発の効率を向上させることが期待される。

バイオマーカーを見つけるためには、生体分子の網羅的な解析が必要となる。DNA (deoxyribonucleic acid) マイクロアレイは、1回の実験で数千から数万の遺伝子発現量を包括的に計測することができ、そのデータはバイオマーカーの探索に利用されてきた。遺伝子発現量はダイナミクスを持つため、複数の時点で計測した、すなわち、時系列のマイクロアレイデータは、単一の時点で計測したマイクロアレイデータと比較して、よりバイオマーカーの探索に貢献できると考えられる。現在、時系列のマイクロアレイデータによるバイオマーカーの探索には次の3つの課題が存在する。

1つ目の課題は、 $p \gg n$ 問題 (p は遺伝子の数、 n は患者の数) と呼ばれる、収集した数千から数万の遺伝子発現量に対してサンプルとなる患者の数が圧倒的に少ないことにより、バイオマーカーの探索が困難になる問題である。特に、時系列のマイクロアレイデータにおいては、バイオマーカーの探索空間に遺伝子数と患者数だけでなく、時系列も加わるため、より探索が複雑で困難になる。

2つ目の課題は、遺伝子発現量の間には存在する多重共線性である。多重共線性とはある遺伝子同士の発現量が高い相関関係にあることを示すが、この相関が誤ったバイオマーカー探索につながる可能性がある。従来の手法では、そのような遺伝子のペアの一方をマイクロアレイデータから削除していた。しかしながら、どちらの遺伝子を削除するのが適切であるかが探索時には明確になっていないため、この方法では重要な遺伝子を削除する可能性があるといった課題があった。

3つ目の課題は、遺伝子発現量のダイナミクスである。同一の患者の生体内であっても、ダイナミクスによって遺伝子発現量は常に変化している。そのため、計測した遺伝子発現量の値を指標とするバイオマーカーでは、計測するタイミングによっては精度が低くなるといった課題があった。

これら3つの課題を解決するために、時系列のマイクロアレイデータから遺伝子発現量を用いたバイオマーカーを探索する2つの手法を開発した。提案手法は、expanding elastic net for consistent differentiation (eENCD) と consolidating probabilities of multiple time points (CPMTP) である。

eENCD は $p \gg n$ 問題と多重共線性の問題を解決するために、機械学習手法の一つである elastic net を採用した。elastic net は、治療効果に関与する遺伝子はごく少数であるという仮定を用いて、膨大な遺伝子数からいくつかの関連遺伝子を選択するスパースモデリング手法である。しかしながら、elastic net は時系列のデータには対応していないため、eENCD は elastic net における遺伝子の重みを活用して、複数の時点で治療効果の有無によって発現量が一貫して異なる値を示す遺伝子を選択できるように拡張した。

CPMTP は、eENCD で取り組んだ 3 つの課題に加えて、eENCD が持つ 2 つの問題に対しても対応している。

その問題の一つは、治療効果の有無によって発現の変化量に大きな違いがあっても、発現量の値のみを見ているために、eENCD ではそのような遺伝子が見落とされる可能性があることである。例えば、治療効果がある患者では 1.0、2.0、3.0 と単調増加の傾向にあるが、治療効果のない患者では 2.0 と一定に遺伝子発現量が維持されている場合である。eENCD は複数の時点で発現量が一貫して異なる遺伝子を探索するため、このような遺伝子を選択できない。そこで、CPMTP は治療効果の有無によって発現量の変化傾向に違いのある遺伝子を探索するために、ある時点と次の時点間における発現量の差分を利用し、発現の変化量に大きな違いがある遺伝子を選択できるようにした。

問題の 2 つ目は、eENCD では時点毎に予測結果が得られ予測結果が複数存在するため、治療戦略を決定することが難しいことである。そこで、CPMTP はベイズの定理を利用することにより、時系列データの各時点で得られた予測結果を統合して 1 つの予測結果を提供できるようにした。

本論文では、実際の多発性硬化症 (MS) と C 型肝炎 (HCV) の患者からマイクロアレイにより収集した遺伝子発現量のデータと治療効果の有無を用いて、eENCD と CPMTP を精度 area under the receiver operating characteristic curve (AUC)、遺伝子の機能の 3 点から従来の手法と評価・比較し、手法の有効性を示した。

本論文は 5 章から構成される。各章の詳細は以下の通りである。

1 章では、本研究の背景と動機について述べた。

2 章では、eENCD の詳細や評価に関して述べた。eENCD は、治療効果の有無などの 2 群間において複数の時点で一貫して発現量が異なる遺伝子を選択する手法である。eENCD は各時点において選択された遺伝子が次の時点においても選択されやすいように elastic net の重みとして組み込むことで、複数の時点で一貫して発現量が異なる遺伝子を選択するように elastic net を拡張した。評価は MS 患者から収集された末梢血単核細胞に由来する 2 つの時系列のマイクロアレイデータを使用して実施された。これらの結果により、eENCD は従来手法より精度と AUC が良く、かつ疾病に関連のある遺伝子を選択できたと考えられ、バイオマーカー探索手法としての eENCD の有効性を示した。

3 章では、CPMTP の詳細や評価に関して述べた。CPMTP は治療効果の有無において発現量の変化量、すなわち発現パターンが異なる遺伝子を選択する手法である。さらに、CPMTP は複数の時点で得られた予測結果を 1 つの結果に統合することが可能である。CPMTP は、各時点での予測に関する情報が多いほど精度が向上するという仮説に基づき、ベイズの定理を応用した。評価は MS 患者と HCV 患者から収集された末梢血単核細胞に由来する時系列のマイクロアレイデータとを使用して実施された。その結果、CPMTP は従来手法より精度と AUC が良く、より早い時点で正確に治療効果を予測できた。また、選択した遺伝子は MS や HCV と関連のあるという報告があった遺伝子であった。これらより、バイオマーカー探索手法としての CPMTP の有効性を示した。また、CPMTP の精度と AUC は、仮説とは異なり、予測に用いる時点が増加しても変化せずに維持された。これにより、治療効果における早期の時点の重要性を示唆した。

4 章では、eENCD と CPMTP の比較について述べた。本比較は MS 患者の時系列のマイクロ

アレイデータを用いて実施された。CPMTP の精度と AUC は、多くの時点で eENCD よりも高かった。また、これら 2 つの手法で選択された遺伝子の機能に対してオントロジー解析したところ、「脂肪酸代謝」、「代謝経路」および「脂質代謝」と MS に関連するものが多かったが、これらに関連している遺伝子は CPMTP によって選択されたものが eENCD によって選択されたものよりも多かった。特に、CPMTP によって選択された *ACCA2* 遺伝子に関連する脂肪酸分解は、先行研究により MS の潜在的な治療標的になる可能性を示唆されていた。したがって、CPMTP は、eENCD よりも疾患に関するより有益な情報を持つ遺伝子を選択できると考えられた。以上より、バイオマーカー探索方法として、CPMTP は eENCD よりも有効である可能性を示唆した。

5 章では、本研究で得られた結論を述べた。

本論文の研究成果により、eENCD と CPMTP は、治療効果を正確に予測し、時系列のマイクロアレイデータの膨大な遺伝子の中から、標的の疾患に関連する遺伝子を選択することができることを示した。これにより、バイオマーカー探索手法としての 2 つの手法の有効性を示した。これらにより選択した遺伝子がバイオマーカーとして臨床応用できるようになるには多くの *in vivo/vitro* での検証が必要であるが、eENCD および CPMTP はバイオマーカーの候補となる遺伝子を提供することにより、個人に合った治療戦略の決定を支援することや、創薬のコストを削減するのに役立つと考えられる。さらに、CPMTP の評価では、治療効果の正確な予測には早期の時点が重要であることを示唆したため、バイオマーカーの *in vivo/vitro* での実験において重点的に評価すべき時点も示唆できたと考えられる。以上より、本研究はバイオマーカーを用いた個別化医療の確立に貢献し、バイオインフォマティクス分野において多大に貢献することが期待できる。