

## 科学研究費助成事業 研究成果報告書

令和 4 年 6 月 7 日現在

機関番号：82606  
研究種目：基盤研究(C)（一般）  
研究期間：2019～2021  
課題番号：19K07744  
研究課題名（和文）線形&非線形分離を考慮したデータマイニングによる新規バイオマーカー同定法の確立  
研究課題名（英文）Development of the screening method for biomarkers to predict the response to dCRT in cancer cells by support vector machine.  
研究代表者  
飯田 直子（Iida, Naoko）  
国立研究開発法人国立がん研究センター・研究所・研究員  
研究者番号：40360557  
交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：バイオマーカー探索に非線形分離手法を利用し、これまで見つからなかった精度の高い新規バイオマーカーを効率よく同定する方法を確立することを目的とした。公開大規模データベースの薬剤感受性データとゲノム網羅的DNAメチル化データを利用し、データの選別、メチル化値データのゲノム位置情報に基づいた圧縮、非線形分離パターンを示すターゲットのフィルタリングを行った後、SVMによるスクリーニングの手法を構築した。データベースを用いた検証では、高精度なマーカーを同定することが出来た。次に、食道がん根治的化学放射線治療奏効性に関するデータを用いた応用を行ったが、検証において、新規バイオマーカーの同定に至らなかった。

## 研究成果の学術的意義や社会的意義

本研究は、バイオマーカーの開発に非線形分離パターンも網羅した点で新しい手法である。細胞株を用いて得られた大規模データベースのデータを用いてスクリーニング方法の構築を行い、本手法で高い精度のターゲットを得られることを示すことができた。しかし、臨床データを用いた応用では、検証において、精度の再現性を示すことは出来なかった。候補プローブを広げた検討により、目的とするメチル化状態を観察することが必要であると考えられる。ターゲットを絞り込む閾値設定が明らかになれば、非線形分離パターンも網羅したSVMにより効率よくバイオマーカーのスクリーニングが可能になると考えられた。

研究成果の概要（英文）：This study aimed to establish an efficient method to identify novel biomarkers with high accuracy that have not been found so far by using nonlinear separation methods for biomarker search. Using drug sensitivity data and genome-wide DNA methylation data from an extensive public database, the methylation value of CpG probes was compressed based on genomic location information, the target probes that showed nonlinear separation patterns for drug sensitivity were selected, and then constructed a screening method using SVM. Validation using the database allowed us to identify markers with high accuracy. Next, we applied the method using data on response to chemoradiotherapy for esophageal cancer. Any novel biomarkers were not identified in the validation.

研究分野：バイオインフォマティクス

キーワード：バイオインフォマティクス がん バイオマーカー

## 1. 研究開始当初の背景

治療に対する効果には個人差がある。適応患者の層別化のためバイオマーカーが必要とされ、探索が進められている。しかし、再現性及び予測精度が高いバイオマーカーを同定することは検体サンプルが十分にあっても難しい。精度を上げるために複数のバイオマーカーを組み合わせる方法も考えられているが、同じ挙動を示す精度の低いバイオマーカーを複数組み合わせても高精度化には限界がある。これまでの手法の問題点として、2群差検定など線形分離を無意識に想定していることがある。その解決には、新たなバイオマーカー同定のアルゴリズムが必要である。

申請者の所属研究室では以前に DNA メチル化アレーデータを用いて、食道がんの根治的化学放射線療法への奏効性を予測する単独バイオマーカーを同定している (Takahashi T et.al. J Cancer Res Clin Oncol, 141: 453-463, 2015)。しかし、実用化のためには、このバイオマーカーの感度と特異度 (39%と 90%) に高精度化が必要である。そこで申請者は、従来の線形分離を想定した組み合わせバイオマーカーの探索を行ってきた。しかし、方法論として線形分離を想定したスクリーニング手法に限界があると考え、本研究でバイオマーカー探索に非線形分離パターンという次の独自のコンセプトを導入し、機械学習によるスクリーニングを実装することにした。

## 2. 研究の目的

バイオマーカー探索に非線形分離手法を利用し、これまで見つからなかった精度の高い新規バイオマーカーを効率よく同定する方法を確立する。まず、既存の大規模データベースのデータを用いてスクリーニングと検証を行い、新規薬剤奏効性バイオマーカーを同定する。これらの解析を踏まえて、汎用性のある効率の良い手法を確立する。次に、確立した手法を食道がん患者の 116 症例の奏効性・遺伝子突然変異・メチル化に応用し、根治的化学放射線治療に対する奏効性予測バイオマーカーの同定を行う。最終的に、新たなコンセプトに基づくバイオマーカーを分離する解析基盤を構築する。

## 3. 研究の方法

まず、手法を確立するためには大規模サンプルデータが必要である。そのために、(1) 公開大規模データベースの薬剤感受性データとゲノム網羅的 DNA メチル化データを利用する。データの選別、メチル化値データのゲノム位置情報に基づいた圧縮、非線形分離パターンを示すターゲットのフィルタリング、SVM によるスクリーニングと検証を行い、バイオマーカーの同定と手法の確立をする。次に、(2) 所属研究室で既に取得している食道がん根治的化学放射線治療奏効性に関する 116 症例のデータ (奏効性、変異、メチル化) を用いた応用を行い、新規バイオマーカーの同定を行う。

### (1) データベースと SVM を利用した線形 & 非線形分離パターンのバイオマーカーの探索

データ選別: 手法を確立するために、近年公開された 2 つの大規模データを利用する。

・1028 がん細胞株の 256 種類の薬剤感受性データ (<http://www.cancerrxgene.org>)

・957 がん細胞株の DNA メチル化データ ([http://cancer.sanger.ac.uk/cell\\_lines](http://cancer.sanger.ac.uk/cell_lines))

実験条件とその結果を分析し、信頼性が高いデータセットを選別する。

データ加工:ゲノムワイドな DNA メチル化は Illumina Infinium Human methylation beadchip 450K により測定されている。DNA メチル化部位(1塩基レベル)はゲノム全体で約 45 万箇所になる。以前に申請者が開発した MACON:DNA メチル化データのプロセッシングツール (<https://epigenome.ncc.go.jp/macon>)を用い、ノーマライズ処理によるメチル化値の算出と DNA メチル化部位のグルーピングによるターゲット数の圧縮を行う。データはスクリーニング用データとバリデーション用データに二分する。

線形 & 非線形分離パターンのためのデータフィルタリング:線形分離に加え、非線形分離パターンを示すバイオマーカーを拾い上げるアルゴリズムを構築する。そのためには、一般的な群の平均値を用いた2群差検定と全く異なる手法が必要である。データの分散、DNA メチル化値についてのヒストグラムによる可視化、ヒストグラムの階級と度数などを算出する。これらの値から基準値を定めて候補ターゲットを選択する。

SVM によるスクリーニングの実行:4つの動きを組み込んだ SVM のプログラムをデザインする。(1)- で取得した各候補ターゲットに対して、(i)4種類の kernel 関数とパラメーターを試行して最適モデルの取得。(ii)5回のクロス・バリデーション。(iii)スコア(正確性、感受性と特異性)の計算。(iv)最適モデルとスコアの出力である。

評価:スコアから候補ターゲットの順位付けをする。さらに、最適モデルとデータの可視化を行って、分離境界線に広いマージンがあるものを上位候補ターゲットとする。

SVM による検証の実行:予め確保しておいた検証用データセットに対して、上位候補ターゲットとその最適モデルを用いて薬剤奏効性予測を行う。

## (2) 検体サンプルを用いた応用

スクリーニング:申請者の所属研究室では既に 116 症例の食道がん根治的放射線治療奏効性情報が有るサンプルを取得し、そのうち 41 症例についてはターゲットシーケンスによる変異解析及び DNA メチル化アレイ解析を完了している。上記で確立した解析手法を活用して予測バイオマーカーのスクリーニングを行い、上位候補ターゲットと最適モデルを同定する。

検証:検証サンプルを用いて、候補ターゲットのメチル化値を Pyro-sequencing 法または MSP 法により測定する。最適モデルを用いた SVM により予測を行い、予測精度を算出する。精度が再現出来れば、新規バイオマーカーとして同定できる。

## 4. 研究成果

2019 年度

データベースと SVM を利用した線形 & 非線形分離パターンのバイオマーカーの探索において、使用するデータの選別とデータフィルタリングを行った。データの選別は IC 値の実験データから、信頼性の高い9個に絞ることができた。さらに CpG island については、長さ 1kb 以上のものについては sub id を作成し、距離を考慮した独自の CpG island の ID を作成した。アノテーション付加を行うスクリプトも作成できた。データフィルタリングにおいて、サンプルのメチル化度合いを評価することが必要であった。そ

のため、TCGAの大規模データを用いて、プローブ毎のメチル化値の分散や平均値などを調べ、メチル化値の分散を用いる評価基準を定めた。この基準を用いて、候補ターゲットのフィルタリングが出来た。

2020 年度

大規模データベースを利用して選別した薬剤の薬剤奏効性とメチル化データを用いて、SVMによるスクリーニングと検証の実行を行い、バイオマーカー同定手法の確立を行った。ワークフローは、(1)ターゲットプローブについて DNA メチル化値データを作成、(2)サンプルの薬剤奏効性データファイルの作成、(3)テスト用データとバリデーション用データの作成、(4)テスト用データを用いたSVMの実行、(5)バリデーション用データを用いたSVMの実行、の5段階を設定し、それぞれのプログラムを作成した。いくつかの薬剤について、スクリーニングを行なった結果、感度、特異度が良い候補ターゲットプローブを得られた。さらに、バリデーションを行い、感度、特異度を保持したターゲットプローブを確認出来た。

2021 年度

116 症例の食道がん根治的放射線治療奏効性情報が有るサンプルを取得し、そのうち 41 症例についてはターゲットシーケンスによる変異解析及び DNA メチル化アレイ解析を行なっている。これらのデータを用い、確立した解析手法を活用して予測バイオマーカーのスクリーニングを行い、上位候補ターゲットと最適モデルを同定した。次に、検証サンプルを用いて、候補ターゲットのメチル化値を Pyro-sequencing 法または MSP 法により測定を行ったが、バイオマーカーとしての精度を再現することが出来なかった。細胞株を用いて得られた大規模データベースのデータを用いてスクリーニング方法の構築を行ったが、臨床データを用いた応用では、検証において、精度の再現性を示すことは出来なかった。ターゲットを絞り込む閾値設定に工夫が必要になると考えられる。候補プローブを広げ、目的とするもののメチル化状態を観察し、さらに精度を上げることが必要だと考えられた。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 飯田直子
2. 発表標題 Detection of alternative splicings and the causative genomic variants with only transcriptome data
3. 学会等名 日本バイオインフォマティクス学会年会
4. 発表年 2020年

1. 発表者名 飯田直子
2. 発表標題 Detecting genomic variants causing abnormal splicing using only transcriptome data
3. 学会等名 日本癌学会学術総会
4. 発表年 2020年

1. 発表者名 飯田直子
2. 発表標題 サポートベクターマシーンを用いたエピジェネティックマーカー同定手法の確立
3. 学会等名 エピジェネティクス研究会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------