

論文 / 著書情報  
Article / Book Information

題目(和文)	疾患と、miRNA配列及びその発現量との関係に関する研究
Title(English)	A study of the relationship between diseases and miRNA sequence with its expression level
著者(和文)	大曾根達則
Author(English)	Tatsunori Osone
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11390号, 授与年月日:2020年3月26日, 学位の種別:課程博士, 審査員:吉田 尚弘,大河内 美奈,芹澤 武,田巻 孝敬,山田 桂太
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11390号, Conferred date:2020/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	要約
Type(English)	Outline

本研究では、最終的にマイクロ RNA (miRNA) 遺伝子の遺伝暗号解読を機械学習で行うための、前処理方法の検討を試みた。

第 1 章では、主に miRNA について紹介する。

第 2 章では、miRNA と生体反応との関係を明らかにするため、これまで miRNA に関して行われてきた研究とは全く異なる手法を試みた。これまでの研究では主にシード理論に基づき、miRNA の標的 mRNA を予測するツールの開発が行われてきた。しかし、シード理論は偽陽性や偽陰性が多いことが明らかになり、近年ではネットワーク解析によって新しい疾患関連 miRNA の予測が行われている。また、機械学習を用いてバイオマーカー探索を行っている研究も少なくない。しかし、本研究が見据えている本研究の先にある「miRNA 遺伝子暗号の解読」という観点において、従来の手法では明らかにすることは困難であると予測される。そこで本研究では、RNA/RNA 相互作用に着目した新たな手法により miRNA を数値化することで、miRNA と疾患の関係を明らかにすることを試みた。そのために、本章では miRNA の数値化に関して検討した。RNA/RNA 相互作用は水素結合を介して行われる。そこで水素結合を行う原子の電荷に基づいた数値化を行った。そのために、ツールを用いて各核酸塩基の三次元構造を構築し、半経験的分子軌道法によって着目している原子の電荷を計算した。数値化においては、電荷の和や電場ベクトルなどいくつかの手法を検討した。その際、miRNA を miRNA の二次元構造と位相幾何学的に相同であるトーラスとみなした。数値化された miRNA に対して、数値化されたことによる一意の値の割合の変化や数値の分布を検証した。その結果、電場ベクトルに基づいた数値化は miRNA 配列と完全に 1 : 1 の関係になることが明らかになった。また、数値の分布は電荷の和のように高い相関を持つものもあれば、電場ベクトルに基づいた数値のように相関がほとんど見られないものもあった。

第 3 章では、第 2 章で数値化された miRNA を用いて、疾患と miRNA との関係を検証した。世界的に死亡者数の多い疾患 4 種類に関して miRNA を用いた研究に関する論文を取得し、論文から各疾患関連 miRNA およびその発現量もしくは発現量比を取得した。疾患の数値化においては、数値化された miRNA とその miRNA の発現量比を掛けたものを用いた。また、論文ごとに発現量比の対数を取るものと取らないものがあるので、どちらか一方に統一し、対数化の有無による比較も行った。また、数値化された疾患は線形性が見られたため線形回帰によって疾患の分類が可能であるかを検証した。さらに、大きな外れ値に対して外れ値を考慮するかしないかによって、どのように分類精度が変化するかも検証した。その結果、対数化を行わず、電場ベクトルに基づいた数値化の一種である EV\_S\_B を用いることで疾患の分類が可能であ

ることが明らかになった。

第4章では、miRNA のシード領域およびシード外領域の寄与率について、また miRNA の検出方法の違いが結果に与える影響について検証を行った。各領域の寄与率においては、各領域に対して第2章の手法により数値化を行った。その後、第3章と同じ操作を、同じデータに対して行った。その結果、第3章で最も適していると考えられた EV\_S\_B では、シード領域のみのスコアにおいても同じ結果を示すことが明らかになった。また、miRNA の検出方法の違いが与える影響について検証を行う際には、第3章で取得した論文データを、miRNA の検出方法ごとに分類し、それぞれに対して第3章と同じ操作を行った。この際、microarray を用いた研究においては用いられたチップが明らかでないものが少なくなかったため、全て異なるチップを利用したとみなした。その結果、次世代シーケンサーを用いるのが望ましいことが示唆された。

第5章では、小分類であるガンの種類が区別できるかを試みた。ガン関連 miRNA とその発現量比に関しては、第3章と同様の手法で収集した。さらに、より正確な物理化学反応に基づいた数値化を行うために水素安定同位体についても考慮した。しかし、従来の量子力学計算ソフトでは最も豊富な同位体のみが考慮されており、水素であれば重水素における電荷計算が行えない。そこで、分子内電荷と関連する電気陰性度の値を用いて重水素の電荷を概算した。この時、電気陰性度が計算可能なオールレッド・ロコウの電気陰性度を用いた。その結果、軽水素のみの miRNA 数値化よりも重水素の存在を考慮した数値化の方がガンの種類が区別できることが明らかになった。さらに、同様の手法でガンのステージも区別できることが示唆された。

第6章では、本研究のまとめと今後の展望を述べる。本研究では、miRNA 配列と疾患には相関があることを示した。しかし機械学習を行うにあたっては十分なデータ数が揃っておらず、また遺伝子であるのにもかかわらず人種の違いが考慮されていない。そこで、日本人に特化したデータベースの構築を提案した。