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)	大曽根達則
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)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： 応用化学 系
Department of, Graduate major in 応用化学 コース

申請学位（専攻分野）： 博士 （ 理学 ）
Academic Degree Requested Doctor of

学生氏名： 大曽根 達則
Student's Name

指導教員（主）： 吉田 尚弘 教授
Academic Supervisor(main)

指導教員（副）：
Academic Supervisor(sub)

要旨（英文 800 語程度）
Thesis Summary (approx.800 English Words)

Because of its function, miRNA can be applied to disease prediction, diagnosis, and therapeutic drugs. The disease prediction/diagnosis can be performed by statistical analysis of miRNA profiles that differ between normal and patient. In addition, the selection of therapeutic agents can be narrowed down by selecting miRNAs that target mRNAs related to the disease as candidates. However, in order to aim for personalized medicine, it is not necessary to use conventional statistical analysis or brute force selection, but to know what each of the 2600 sequence patterns in the miRNA gene means. is important. In addition, since miRNA is a functional RNA, decoding the miRNA gene code may also reveal the meaning of other functional RNAs. Furthermore, the miRNA gene code can be deciphered to enable the prediction of RNA that would have some function in the RNA world on the primitive earth. It is also suggested that it leads to the origin of life.

The purpose of this study was to clarify the relationship between miRNA sequences and diseases, and to be a step toward deciphering the miRNA gene code. In conventional research, seed theory and network analysis have been used. In this study, we clarified the relationship between sequences and diseases by scoring miRNAs focusing on the hydrogen-bonding atoms used in RNA / RNA interactions. As a result, it was suggested that there was a high correlation between the score and the disease, and that the disease could be classified. Since scoring and miRNA sequences in this study are approximately 1: 1, it is thought that there is a correlation between miRNA sequences and diseases.

There are two main problems in this study. The first is that the number of data including the number of target diseases is small, and the second is that it is not yet clear what the score means.

For cases where the first number of data is small, we would like to conduct a large-scale screening for Japanese people and to conduct analysis specialized for Japanese people because we would like to eventually connect to personalized medicine. In particular, cancer data has high social significance and it is considered that samples are easy to collect.Thus, in collaboration with university hospitals and cancer centers, a large amount of short RNA sequences of Japanese sequences were obtained. I want to apply the method. The reason for specializing in the Japanese is that miRNA is a gene, so the profile may differ between races.

I would like to do machine learning for the case where the meaning of the second score is not clear yet. We did not use machine learning directly in this study because it is difficult to machine-learn the relationship between miRNA sequences themselves and diseases. Various methods are conceivable as a machine learning method. As a simple method, a method of performing language processing by regarding the miRNA sequence as a character string sequence (or sentence) can be considered. However, this method is difficult from the following two points. 1) RNA has only 4 types of character elements. 2) miRNA is only about 20 bases long. Therefore, it is necessary to use a different method. For example, in the case of deep learning, it is necessary to provide appropriate parameters and inputs in advance. Taking image discrimination as an example, the input extracts features such as contour lines at various resolutions. Currently, this method has been established, and libraries (which differ depending on programming languages such as packages) have been created, so anyone can easily perform deep learning of image recognition. However, if the method described above is not input, the RGB ratio is input, or the RGB position pattern is regarded as a matrix and some linear processing is input, It is considered that an algorithm with such a high discriminatory power could not be obtained. Therefore, in order to apply the relationship between miRNA sequences and diseases to machine learning, it is necessary to search for appropriate parameters and inputs. The findings obtained in this study suggest that quantifying RNA sequences as charges may be suitable as an input for machine learning. In addition, it is predicted that the expression level of miRNA may be used as a parameter.