

令和 6 年 5 月 28 日現在

機関番号：40115

研究種目：挑戦的研究（萌芽）

研究期間：2021～2023

課題番号：21K19238

研究課題名（和文）長鎖シーケンシング技術を利用した環状RNAの全長一発同定手法の開発

研究課題名（英文）Development of a one-shot identification method for full-length circular RNAs using long-sequencing technology

研究代表者

梅影 創（Umekage, So）

函館短期大学・食物栄養学科・講師（移行）

研究者番号：30419436

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：環状RNAは新たなバイオマーカーとして期待されており、新たな環状RNAの探索が精力的に進められている。従来の環状RNAの探索手法は、ショートリードシーケンシング技術をベースとしており、環状RNAの全配列を一気に決定することができないという欠点があった。この研究では、この問題点を克服することを目的として、ロングリードシーケンシング技術に着目して、環状RNA配列の全長を一気に決定できる手法の開発を目指したプログラムを作成した。一方、環状RNA由来のコンカテマー配列を作成するロングリードシーケンシングライブラリ作成方法を完成させることができず、今後の課題となった。

研究成果の学術的意義や社会的意義

環状RNAをロングリードシーケンシングによって一気に解決することができれば、バイオマーカーとなりうる環状RNAの探索がよりスムーズになると期待できる。また、検体試料（例えば唾液）から環状RNA検出までの全行程を確立することができれば、新たな健康診断ツールとしての利用が期待される。

研究成果の概要（英文）：Circular RNA is expected to be new biomarkers, and the search for new cyclic RNA for diagnostic uses is being vigorously pursued. Conventional circular RNA finding method is mainly based on short-read sequencing technology, which has the disadvantage that the entire sequence of circular RNAs cannot be determined at once. In this study, I focused on long-read sequencing technology to overcome this problem and aimed to develop a method that can determine the entire length of cyclic RNA sequences at once. A Python program was created to exclude biased sequences from the long-read sequencing data and to detect circular RNA sequence using a string matching library, RapidFuzz. On the other hand, the method for creating a long-read sequencing library to generate concatemers derived from circular RNAs was remained incomplete.

研究分野：RNA biology

キーワード：circular RNA

### 1. 研究開始当初の背景

環状 RNA は末端部分を持たない構造をしており、エキソ型ヌクレアーゼによる分解を受けないため、直鎖状 RNA と比べて細胞内環境で安定に存在することができる (図 1)。また、細胞の老化や癌化などにおいて環状 RNA の発現量が変化することから、環状 RNA は新たなバイオマーカーになりうるものと期待され、新規な環状 RNA 探索が精力的に行われている。環状 RNA の探索方法については、ショートリードシーケンシング技術を利用し、環状化した RNA 配列部分を検出するという方法であり、環状 RNA の全長を決定できないという不便さがあった。

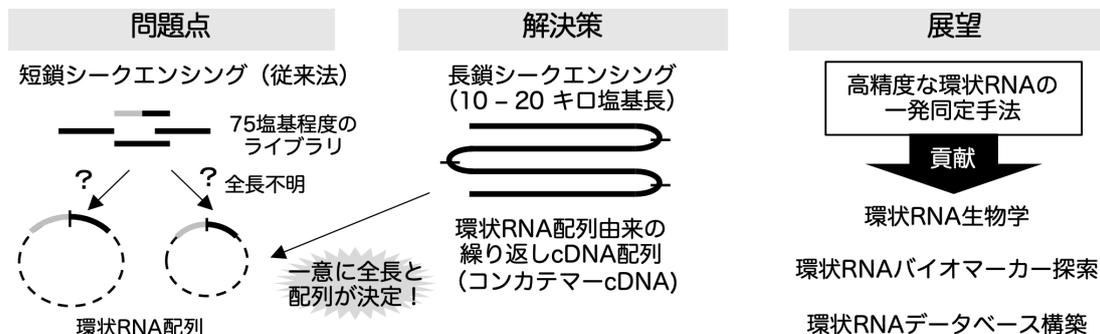


図1 本研究の概要

### 2. 研究の目的

本研究では、環状 RNA 全長の高精度一発同定手法の確立を目的として、10 kb~20 kb の平均リード長である Oxford nanopore technology 社のナノポアシーケンサーを用いた長鎖シーケンシング技術を活用し、実験面では、(1)環状 RNA の全長を決定するための環状 RNA 由来コンカテマー (繰り返し配列) cDNA ライブラリ調製方法の開発を、解析面では、(2)コンカテマー配列の特徴を活かした環状 RNA 配列を高精度に同定する手法の開発を実施する。また、本手法の実用性の検証として、(3)本手法を用いた新規環状 RNA の探索を実施する。環状 RNA 生物学研究の進展だけでなく、臨床応用を目指した環状 RNA バイオマーカーの発見、環状 RNA データベースの構築などへの貢献を目指す。

### 3. 研究の方法

シーケンシングによって得られたデータから環状 RNA の配列を決定するための情報科学的手法と環状 RNA 配列の検出を目的としたロングリードシーケンシング用ライブラリの作成について検討することとした。

#### 3-1 手法の開発について

- fastq ファイルの判定 (フォーマット確認) 用 Python プログラムの作成
- 冗長な配列、N の多い配列、繰り返し配列判定用 Python プログラムの作成
- 環状 RNA の判定を妨げる短い配列除去用 Python プログラムの作成
- コンカテマー配列を検出するプログラムの作成
  - 1 ゲシュタルトパターンマッチング DiffLib の検討
  - 2 レーベンシュタイン距離を計算する RapidFuzz の検討

#### 3-2 環状 RNA 配列の検出を目的としたロングリードシーケンシング用ライブラリの作成について

実施不足のため省略

### 4. 研究成果

ロングリード配列から生成する繰り返し配列 (コンカテマーと呼ぶ) を情報科学的に検出する手法として 2 つの文字列の差分を計算する diffLib ライブラリの使用を検討したが、コンカテマー検出に長時間を要するため実用的でないことがわかった。ついで、文字列の類似度を高速に計算する RapidFuzz を利用したところ、検出時間の問題が解決された。一方、1 文字や 2 文字の繰り返し配列を後検出する傾向がみられたため、RapidFuzz プログラムの実施前に、上記繰り返し配列を検出するプログラムを新たに組み込むことでこれらの問題が完全ではないものの一応



5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔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
---------	---------