

令和 4 年 6 月 10 日現在

機関番号：12602

研究種目：基盤研究(B)（一般）

研究期間：2019～2021

課題番号：19H03214

研究課題名（和文）逆転写酵素による鋳型非依存的な塩基付与の法則の理解と完全長cDNA合成の精緻化

研究課題名（英文）Improvement of cDNA conversion efficiency to detect full-length sequences of extremely low amounts of RNA

研究代表者

笹川 洋平（SASAGAWA, YOHEI）

東京医科歯科大学・難治疾患研究所・准教授

研究者番号：10404344

交付決定額（研究期間全体）：（直接経費） 12,800,000円

研究成果の概要（和文）：本研究では、1細胞以下の細胞内区画や細胞小器官などの検体に存在する超微量RNAの全長配列を高感度に捉えるために、cDNA変換の精緻化及びcDNA全長配列のハイスループットな検出方法の開発を行った。cDNA変換の精緻化では、効率を改善させる複数の候補因子を発見した。加えて、長鎖シーケンサーに適した全長cDNA検出方法への最適化も行い、おおむねのワークフローを確立することに成功した。

研究成果の学術的意義や社会的意義

1細胞以下の超微量RNAの検出の要望は高まっておりcDNA変換効率の改善が必要であるが、ほとんど改善されないまま使われていた。またcDNA変換は、基礎的な分子生物学的な技術であり、その改善は、微量RNAを扱う研究者・産業界など広範囲に影響を及ぼすと考えられる。本研究での知見は、幅広い分野に直接的な技術貢献をするとともに、超微量RNAからの完全長RNAを可能にし、未知RNAの発見につながると期待される。

研究成果の概要（英文）：The project's aim is to develop a highly sensitive method for detecting full-length sequences of extremely low amounts of RNA present in intracellular organelles and other samples of less than single-cell. Therefore, in this study, I attempted to improve cDNA conversion efficiency and the sequencing method to detect full-length cDNA sequence. As a result, I discovered several candidate factors that improve the efficiency of cDNA conversion. I also succeeded in establishing a workflow to efficiently analyze amplified cDNA on a long-read sequencer.

研究分野：ゲノミクス

キーワード：cDNA変換 長鎖RNA配列

## 1. 研究開始当初の背景

(1) 本研究開始に至る申請当初(2018年当時)複数の1細胞RNAシーケンス法が開発・報告されており1細胞レベルのRNAの検出が世界中で行われるようになりつつあった。一方で、1細胞以下の細胞内区画や細胞小器官ならびに細胞外に放出される膜小胞などの検体に存在する超微量RNAからの全長RNA配列の検出の要望が高まっていたが、感度や検出精度などに問題があり困難であった。図1に示す、Whole-transcript amplification (WTA) と呼ばれる分子生物学的な基礎反応が、1細胞以下のRNAを対象とするRNAシーケンスの手法ごとに異なっており、感度の違いはWTA、特にcDNA変換効率に依存していると推測された。そのため、cDNA変換の抜本的な改善が必要と考えられた。また上記微量RNAシーケンス手法では、短鎖シーケンサーによるRNA配列の一部のみを読み取りが一般的なため、一本のつながったcDNA配列として読み取りできなかった。このように、超微量RNAからの全長RNA配列取得のためには、RNA検出感度の上昇と長鎖シーケンサーを用いた全長cDNA配列取得技術の確立の2つが必要であった。

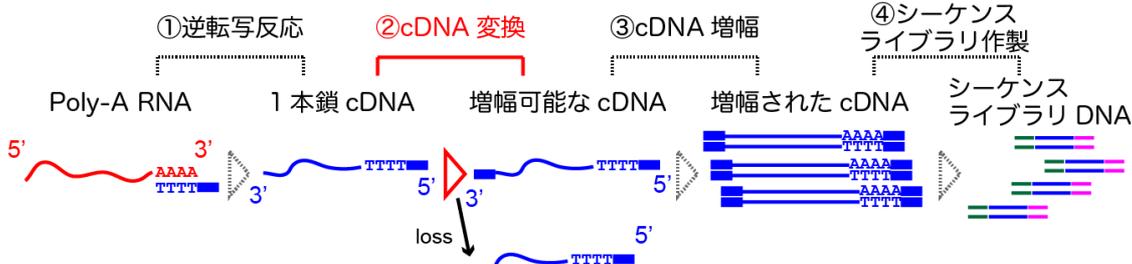


図1. 微量RNAから配列を検出するための分子生物学上の必須過程  
Whole-transcript-amplification (WTA)は、の工程を含む。

## 2. 研究の目的

### (1)

本研究では、1細胞以下の細胞内区画や細胞小器官ならび細胞外に放出される膜小胞などの検体に存在する超微量RNAの全長配列を検出するため、上記背景で述べたcDNA変換の特性の把握と精緻化を目標の一つとした。cDNA変換の精緻化として、poly-A taggingとtemplate switchingの2手法を対象とした。特に、template-switchingに関しては、メカニズムが不明な点にも焦点を当て研究を進めた。また、鋳型依存的にまた超微量RNAの全長配列を検出するために、増幅したcDNAを長鎖DNAシーケンサーによる効率的に読み取る手法の確立も目指した。これら2つの主項目について、以下の研究方法と研究成果に分けて記載する。

## 3. 研究の方法

### (1)

qPCRによるcDNA変換効率の計測方法について記す。ポリA配列を3'末端に持つ、100nt以下の短い人工RNAを用意し、5'末端の数塩基をランダム配列にした。アダプター配列を持つoligo-dT配列を用いて逆転写反応を行い、1本鎖cDNAを合成し、poly-A taggingやtemplate-switching反応を行い、もう片側にアダプターを付与する。cDNA変換が行われると、両端のアダプターでPCR増幅できる。これを利用して、cDNA変換の効率を計測した。

### (2)

次世代シーケンス解析によるpoly-A RNAの検出方法。Quartz-Seq2のWTA及びsequence library DNA作製手法をベースに、WTAの部分を改変し、sequence library DNAを作製した(引用文献#1)。解析に関しても、引用文献#1を改変し、発現定量を行った。Cell-barcode配列やMolecular-barcode配列はそのまま使用し、transcript sequenceでは、人工RNAではその配列を使用し、各生物種のRNAを使用する場合は、それぞれの生物種のゲノム配列などを使用し発現定量を行った。次世代シーケンサーには、Illumina社のNextSeq500を使用した。

### (3)

Poly-A taggingの精緻化では、Quartz-Seq2の条件をベースに、条件を追加・改変する実験を進めた(引用文献#1)。qPCRでは、SybrGreenなどの蛍光色素をPCR溶液に加えリアルタイムPCR装置で計測した。

### (4)

バーコード配列の振り分けには、Flexbarを使用した(引用文献#2)。コードは、<https://github.com/seqan/flexbar>を参照。

## 4. 研究成果

### (1) template-switching 反応のメカニズム

本研究では、超微量な RNA からの全長配列を取得するために、精緻化された高感度な cDNA 変換手法の確立を試みた。そのために 2 種類ある cDNA 変換手法 (poly-A tagging と template-switching) の両方をベースに精緻化を試みた。特に template-switching 手法に関しては、1 本鎖 cDNA の 3' 末端への塩基付与のパターンやメカニズムが不明であった。そこで人工 RNA を用いた template-switching 効率の計測法や次世代シーケンスでの検出方法を確立した。一方で、上記メカニズム解明を本格的にはじめた矢先に、template-switching のメカニズムを詳細に調べた論文が、New England Biolabs 社から報告された (引用文献#3)。

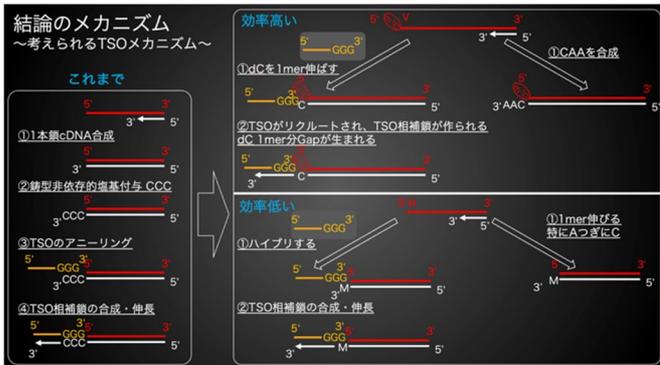


図 2. 引用文献#3 の内容から筆者がまとめた template-switching 反応のメカニズムの模式図。左は、これまで考えられていた template-switching 反応のメカニズム。右は、文献#3 から示唆されたメカニズム。Cap 構造は、m<sup>7</sup>G として表現されている。

本論文による結果だと、template-switching 反応は、鋳型非依存的に C 塩基が一つ付与されたあと、確率的に TSO がうまく入り込んだ場合のみ、進む反応であることを強く示唆していた (図 2 参照)。これらの実験は、非常に短い人工的な RNA の場合で、長鎖 RNA の場合は、更に完全長 cDNA が作られる確率を掛け算するので更に低くなると考えられた。

### (2) 1 細胞 RNA シーケンス手法の再解析

Template switching で最も性能が高いと主張する 1 細胞 RNA シーケンス法である、Smart-seq3 (引用文献#4) を再解析することで、template-switching を用いた RNA シーケンス法の感度の現状を把握することにした。既に報告のあった、我々のチームを含む 13 の手法を公平に比較した、1 細胞 RNA シーケンス手法の性能比較研究のデータと合わせて解析した (引用文献#5)。図 4 に示すとおり、template-switching を使用した Smart-seq2 や Smart-seq3 は 2000 遺伝子前後であった。一方で、template-switching を使わない手法、CEL-seq2 や Quartz-Seq2 では 3000 遺伝子を超えていた (図 3 参照)。他の細胞型においても同様の傾向が観察された。これらの結果は、現状においても template-switching 手法は、感度に伸び悩んでいることを示唆していた。

これらの結果から、Template-switching 反応は、反応原理上これ以上の高感度化が難しく、同反応を使った手法の感度の結果もそれを裏付けていた。本研究の期間内で、template-switching をベースに感度の精緻化を達成するのは困難だと判断し、poly-A tagging 法をベースにさらに高効率の cDNA 変換手法を開発することにした。

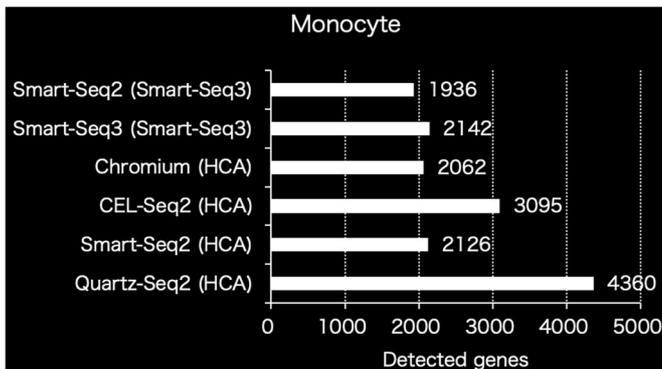


図 3. 引用文献#4 及び#5 からの 1 細胞 RNA シーケンス法の再解析結果。細胞型を分離したあと、Monocyte における検出遺伝子を比較。Y 軸は、手法を示し、カッコ内は文献 11 由来が Smart-Seq3、文献 12 由来が HCA と記載。X 軸は、1 細胞あたり 2 万リードのデータ量あたりに検出される遺伝子数となる。

### (3) poly-A tagging の簡素化

Poly-A tagging の精緻化には、我々が過去に開発した世界最高精度の 1 細胞 RNA シーケンス法 Quartz-Seq2 をベースとした (引用文献#1, #5)。本手法では、terminal transferase によって poly-A を付与したあと、DNA polymerase により 2 本鎖 cDNA を合成し、その後 PCR 増幅する。こちら、後半の 2 本鎖 cDNA 合成と PCR 増幅反応を、感度を下げないまま 1 工程で行えるように簡素化した。これにより、poly-A tagging 反応効率を改善する因子を qPCR によりスクリーニングできる手立てを整えた。

### (4) poly-A tagging の精緻化

上記簡素化により、384well plate をつかって、qPCR により poly-A tagging 効率を計測できるようになった。そこで、導入した機械なども活用し、poly-A tagging の様々な条件を用意してスクリーニングを行った (図 4 参照)。

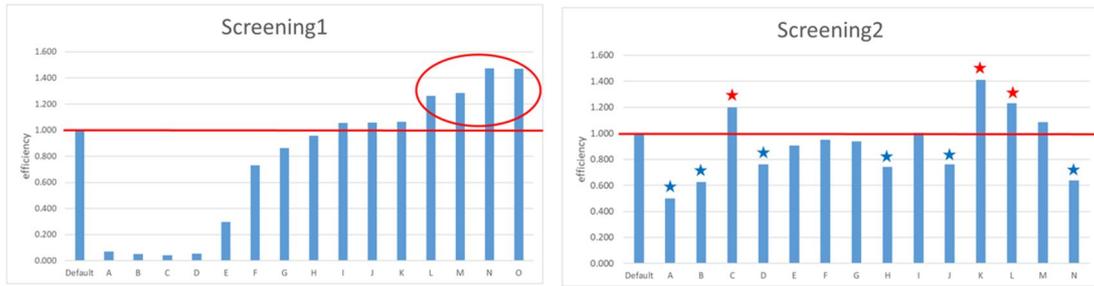


図 4. 様々な poly-A tagging の条件で、その効率を計測した棒グラフ  
X 軸は、各スクリーニングにおける条件名で、Default は通常条件。赤丸や、赤星に示した条件は、通常条件より効率が高かった。赤線は、通常条件での効率を示す。

その結果、いくつもの改善する候補因子をスクリーニングで発見できた。また poly-A tagging に使用する緩衝液をスクリーニングすることで増幅 cDNA 長が改善する条件があることもわかった。これらの条件は、組み合わせることが可能であり、現在ブラッシュアップして最適な組み合わせを検討している。

#### (5) 長鎖シーケンス仕様への変換の試み

長鎖シーケンスに適した増幅 cDNA 構造をつぎに検討した。理想的には、cDNA の両端には別々の配列が付与されることが望ましい。まず逆転写プライマーに使用するアダプター配列と、cDNA 変換に使用するアダプター配列の 2 種類が cDNA の両端につくことで PCR 増幅できるが、これらのアダプター配列の種類のシーケンスに及ぼす影響を調べた。使用するアダプター配列によって、transcript 由来のシーケンスリードに特徴的な塩基パターンが見られた。Transcript 由来のシーケンスリードに 10 塩基以下での塩基組成の乱れは、断片化の処理などの影響によると考えられたが、それ以上の長さの塩基組成の乱れは、アダプター配列の混入だと考えられた。Quartz-Seq2 のアダプター条件では、そのようなアダプター由来の塩基乱れのパターンは見られなかったため、M 配列をアダプターとする逆転写プライマー条件をベースに次に進めた。

#### (6) アダプター配列の選定

両端のアダプター配列を異にするため、poly-A tagging で使用するアダプター配列を複数用意して、検討した。Genome に mapping しないなど、複数のパラメーターから計算されたアダプター配列 8 種類を用意して、PCR 増幅における primer dimer などの副産物への影響を見た（図 5 参照）。通常条件では、byproduct は殆ど出ないが、配列によって byproduct がたくさん出てきた。この中で、byproduct が出てきにくい配列を 2 つ見出した。また同アダプターが、RNA からの cDNA 増幅に使用できることを別途確認した。



図 5. Quartz-Seq2 で増幅される WTA cDNA の構造とその改変  
左上は Quartz-Seq2 の通常条件で増幅される cDNA の構造。左下は、紫の Tagging primer 部分  
を設計した配列に置きかえた。右は、これらの配列を使って、鋳型 RNA を入れないで、増幅反応  
した際の産物を電気泳動で解析し byproduct を観察した。

#### (7) 長鎖シーケンス解析一般

Quartz-Seq2 の増幅 cDNA が長鎖シーケンサーで読み取れるか、Oxford Nanopore Technologie 社の GridION R9.4 および Q20+キットで解析を行った。R9.4 では、Q score が 10 程度であり、シーケンスの quality として 10 塩基に 1 塩基のエラー率であった。一方で、Q20+キットで解析した結果、Q score は 13.4-14 で、おおよそ 22-25 塩基に 1 塩基のエラーであった。このため、新しい解析キットだと、2.2-2.5 倍の品質が向上した。Q20+キットのデータから塩基決定を再度行い、高精度設定で再計算を行った。通常と比べて、約 5 倍の計算時間がかかるが、Q score は 16.3-16.7 まで改善した。これは、43-47 塩基に 1 塩基のエラーの頻度となる。また高精度設定

のほうが、通常設定に比べて 1.2-1.4 倍ほどリード数が多くなった。これらの結果から、高精度設定による Q20+キットによる長鎖 DNA 解析は有用であると考えられた。一方、高精度設定における計算負荷は非常に大きく今後の課題になりそうであった。次に、シーケンス解析された長さの解析を行った結果、8kbp ほどまでは解析できていた。データベース上での Transcript の大部分を含む長さであった。

#### (8) 長鎖シーケンス解析からの demultiplex

解析した増幅 cDNA は 384 検体由来の total RNA となり、それぞれ 14mer のバーコード配列で標識されており、バーコード配列を読み取ることで、どの検体由来なのか判別することができる設計になっている。Flexbar を用いて、バーコードの振り分け率を計算した(引用文献#2)。同 cDNA を Illumina 社の短鎖シーケンサーで解析すると、Q score が 30 以上の塩基が 90%以上であり、バーコードの認識率は、95%以上であった(図 6 参照)。長鎖シーケンサーの R9.4 の場合、完全一致のバーコード認識率は約 10%であった。一方で、高精度設定の Q20+キットだと、完全一致のバーコード認識率は、約 40%であった。このように、長鎖シーケンスの品質が向上することにより、沢山の検体を解析する際のバーコード認識率の改善につながった。今回使用しているバーコード配列は、シーケンスレーベンシュタイン距離による補正を考慮した設計で、具体的には、置換、欠失、挿入などの塩基配列エラーを 2 塩基分補正することができる。こちらを使い、認識率の低い R9.4 のバーコードの認識が補正できるか確かめた(図 6 参照)。

		Mean read length:	Median read length:	Number of reads:(合計)	(Number of reads比率(合計))	Total bases(合計)	Total bases比率(合計)	
ONT	元ファイル	191018_GridION_IP013-24_orig	1660	1430	7686841		12760335803	
	demultiplex(m=0)	191018_GridION_IP013-24_demultiplex_m=0	2110.732031	1957.549479	1611158	20.9599%	3508653382	27.4966%
	demultiplex(m=1)	191018_GridION_IP013-24_demultiplex_m=1	2233.252865	2079.089844	3530702	45.9318%	7630629128	59.7996%
	flexbar(m=0)	191018_GridION_IP013-24_flexbar_m=0	2062.707031	1911.321615	784431	10.2049%	1668484502	13.0756%
	flexbar(m=1)	191018_GridION_IP013-24_flexbar_m=1	2060.009635	1909.007813	799150	10.3963%	1696564936	13.2956%
	flexbar(m=2)	191018_GridION_IP013-24_flexbar_m=2	1974.06224	1814.269531	6237929	81.1507%	11567955711	90.6556%
illumina	元ファイル	RT625_96hRNA3ngV31_p7_S1_ds4595840_R1_orig	22	22	45,958,400		1,011,084,800.00	
	元ファイル	RT625_96hRNA3ngV31_p7_S1_ds4595840_R2_orig	64	64	45,958,400		2,941,337,600.00	
	flexbar(m=0)	RT625_96hRNA3ngV31_p7_S1_ds4595840_R1_m=0	22	22	43,814,711.00	95.3356%	963,923,642.00	95.3356%
	flexbar(m=0)	RT625_96hRNA3ngV31_p7_S1_ds4595840_R2_m=0	64	64	43,814,711.00	95.3356%	2,804,141,504.00	95.3356%

図 6 R9.4 による長鎖シーケンス結果のバーコード振り分け効率

#m=0 はバーコード補正がない場合で、m=1,2 はバーコード補正を 1 もしくは 2 行った場合の結果。ONT は、R9.4 で解析したシーケンス結果のバーコード振り分けをした結果で、Illumina は NextSeq500 で解析したデータをバーコード振り分けした結果になる。

R9.4 で、10 塩基に 1 塩基エラーのシーケンスの品質の場合、エラー補正が 1 だと、ほとんど振り分け率は変わらなかった。一方で、エラー補正 2 の場合、大幅に改善し 70%以上が振り分け可能になった。高精度設定の Q20+のデータは、エラー補正 0 で 40%ほど振り分けができ、エラー補正有りですれほど改善できるか引き続き調査中である。

#### (9) まとめ

本研究では、cDNA 変換の特性の把握と精緻化、増幅した cDNA を長鎖 DNA シーケンサーで効率的に読み取る手法の確立の 2 つを目指した。精緻化においては、poly-A tagging のスクリーニングを用いた手法により改善する因子を複数見出しさらなる高感度化に目処をつけた。また長鎖シーケンサーでの読み取りも可能になった。このなかで、特にシーケンス品質による影響が大きいものの、計算手法を組み合わせることで、大幅な改善が得られた。これら 2 つを組み合わせ、超微量 RNA から高精度に長鎖 cDNA シーケンスで検出するまでのワークフローが確立できた。

#### <引用文献>

- #1 <https://doi.org/10.1186/s13059-018-1407-3>
- #2 <https://doi.org/10.1093/bioinformatics/btx330>
- #3 <https://doi.org/10.1074/jbc.RA119.010676>
- #4 <https://doi.org/10.1038/s41587-020-0497-0>
- #5 <https://doi.org/10.1038/s41587-020-0469-4>

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計5件（うち招待講演 5件 / うち国際学会 0件）

1. 発表者名 笹川 洋平
2. 発表標題 実験技術開発におけるデジタル化の現状と取り組み
3. 学会等名 ジャパンリンクセンター「対話・共創の場」（第7回）コロナ禍を背景とした研究のデジタル化ソリューションに向けて（招待講演）
4. 発表年 2021年

1. 発表者名 笹川 洋平
2. 発表標題 実験技術開発における研究データ公開の役割について
3. 学会等名 2020年度第1回J-STAGEセミナー「ジャーナルから見た研究データ：研究データ公開の意義（招待講演）」
4. 発表年 2020年

1. 発表者名 笹川 洋平
2. 発表標題 1細胞RNAシーケンス法における反応原理の理解に基づくデータの評価方法の紹介
3. 学会等名 シングルセルゲノム研究会2019（招待講演）
4. 発表年 2019年

1. 発表者名 笹川 洋平
2. 発表標題 高出力型1細胞RNAシーケンス法Quartz-Seq2の開発と最新の技術動向
3. 学会等名 第60回日本生化学会中国・四国支部例会（招待講演）
4. 発表年 2019年

1. 発表者名 笹川 洋平
2. 発表標題 1細胞RNAシーケンス法Quartz-Seq2の開発と表現型スクリーニング法への応用
3. 学会等名 ギルソン社ピペットマン生誕50周年記念 研究フォーラム（招待講演）
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 二階堂 愛, 笹川 洋平, 團野 宏樹	4. 発行年 2020年
2. 出版社 化学同人	5. 総ページ数 183
3. 書名 シングルセル解析でなにがわかるか (DOJIN BIOSCIENCE SERIES) Part2 シングルセルでみる生命】4章 シングルセルトランスクリプトーム)	

〔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
---------	---------