

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 22 日現在

機関番号：82401

研究種目：挑戦的萌芽研究

研究期間：2015～2016

課題番号：15K14425

研究課題名(和文) 一塩基同定・一分子定量精度をもつゲノムワイド非コードRNAデジタル計数法の開発

研究課題名(英文) Development of genome-wide digital quantification method for non-coding RNAs which has single-base identification and single-molecule level quantification

研究代表者

城口 克之 (SHIROGUCHI, KATSUYUKI)

国立研究開発法人理化学研究所・生命システム研究センター・ユニットリーダー

研究者番号：00454059

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：分子バーコードの機能を拡張させ、非コードRNAも測定可能とし、一塩基の違いを区別して1分子の精度で定量する方法の開発を目的とした。また、本技術の世界的な標準化も目指した。同じバーコード配列を持つターゲット配列は同じ分子から増幅されたということを利用し、多数決の原理でエラーを同定し、一塩基の精度で配列を同定した。本方法は、核酸を対象にした様々な測定に利用されていくと考えられる。

研究成果の概要(英文)： Our goal was to develop a new digital RNA sequencing which could be applied for non-polyA tailed RNA and had single base resolution. Then, we aim to make a standardized high throughput digital RNA sequencing for researcher all over the world.

In this study, we attached one of many different molecular barcodes to each target molecules stochastically, amplified these molecules, and sequenced them. By counting the number of different barcodes, we quantified the original (i.e., before amplification) number of target molecules. Since the sequenced target molecules that have the same barcode sequence are from the same original molecules, we identified errors, and finally, we determined target sequences with single base resolution. I think this method will be used for quantification of nucleic acid molecules in many measurements.

研究分野：生物物理、定量ジェノミクス

キーワード：シーケンシング デジタル計測 バーコード

1. 研究開始当初の背景

次世代シーケンサが発展し、RNA シーケンシングにより、定量的な網羅的遺伝子発現解析が行われるようになっていた。その数年後に申請者らを含めて分子バーコード法が開発され、より正確に発現量 (RNA の量) を定量できることになった。この方法は、RNA の絶対量が少ない 1 細胞解析を行う場合を含め、世界の多くの研究グループが用いるようになってきていた。一方で、分子バーコード法はその扱い方に注意が必要であり、しっかりと理解して使用しないと、バーコードのデザインや解析法が測定結果に影響してしまうことが分かりつつあった。しかしながら、バーコードの使用や解析法について、統一的な見解が得られていなかった。このような状況の中で、世界の誰もが正確に分子バーコード法を使用できるようになることを目指して、我々は、分子バーコード法の標準化を行ってきた。具体的には、より正確な計測を実現しながらも低コスト化が可能なバーコードのデザイン、そのデザインに応じた解析法を開発してきた。同時に、これまでにない高いダイナミックレンジを実現できることを実験的に示してきた。このように分子バーコード法を発展させてきたが、さらに新しい機能を開拓できるポテンシャルがあると考え、本研究では、その機能の実装とさらなる標準化を目指した。

2. 研究の目的

本研究では、分子バーコードの機能をより拡張させ、非コード RNA も測定可能とすること、また、一塩基の違いを区別できる塩基同定精度を基に、一分子の分解能で定量計測することを目指した。さらに、これらの定量測定をハイスループットに行うシステムの構築も目的とした。また、誰もが使用できるようにするため、解析法をパイプライン化し、世界における標準化の実現を意識した。

3. 研究の方法

(1) デジタル RNA シーケンシング

RNA シーケンシングでは、まず逆転写酵素を用いて RNA と相補的な配列を持つ cDNA を作製する。この cDNA を増幅し、増幅産物の数を次世代シーケンサにより計数する。デジタル RNA シーケンシングでは、増幅前の cDNA 分子それぞれに異なる分子バーコード (DNA 配列) を付加する。増幅後の分子バーコード付 cDNA をシーケンシングするが、この時、増幅産物の数を数えるのではなく、分子バーコードの種類数を計数する。この方法により、増幅前の分子の数 (絶対値) を定量することができる。この方法は、最終的な測定値 (分子の数) が各分子の増幅率に依存しないため、測定の実現性が高く、各遺伝子 (RNA) 間の発現量を比較できるという利点がある。本研究ではこのデジタル RNA シーケンシング法を用いた。

(2) 非コードな短い RNA への分子バーコードの付加

通常の RNA シーケンシングの対象は、3' 末端に polyA が付加された RNA (主に mRNA) が主である。この時、cDNA を合成する際に polyT プライマーが使用される。本研究では、polyA が付加していない RNA (マイクロ RNA などの非コード RNA) の定量も実現するために、まずアダプター配列を 20 ヌクレオチド程度の RNA に付加した。ターゲットとして、購入した合成 RNA を用い、付加する配列として、3' 端にバーコード配列と増幅用の配列を含んだ一本鎖 DNA を用いた。一本鎖の RNA に一本鎖の DNA を付加できる特性をもつ RNA リガーゼという酵素を用い、酵素反応でバーコード配列の付加を行った。

(3) 一塩基でターゲット配列を同定する方法

まず、たくさんの種類のバーコード配列を準備し、目的の分子（ターゲット配列）に確率的に付加した。これにより、統計的に、それぞれのターゲット配列は異なるバーコード配列をもつことになる。増幅後のサンプルの配列を、次世代シーケンサで同定した。得られた配列を解析して、一塩基の精度でターゲット配列を同定する方法の開発を目指した。まず、共同開発したソフトウェアを用いて、得られた配列をバーコード配列毎にクラスター化した。この時、同じクラスターとして分けられたターゲット配列は、同じ分子由来の増幅産物であることを示している。同じクラスター内のターゲット配列を、既存のソフトウェアを用いてアライメントし、共通の塩基、また、異なる塩基を同定した。異なる塩基については、多数決の原理に基づいて、一番多い塩基が正しい塩基とした。尚、これらの解析用にプログラムを作製した。増幅時のエラーの影響を低減するために、リニア増幅にもトライした。

4. 研究成果

(1) 非コードなマイクロRNAへの分子バーコードの付加

20塩基程度の合成RNAに、バーコード配列と増幅用のプライマー配列を含む一本鎖DNAを付加した。電気泳動法で確認したところ、バンドの移動が起きており、およそ50%の合成RNAにバーコード配列を付加することができたと考えられる。50%は高い効率と言えるが、さらに効率を上げられる可能性もあるので、条件を検討していく。

(2) 一塩基の同定精度

増幅時のエラーやシーケンシングでのエラーがまったくなければ、増幅後に同じバーコードを持っているターゲット配列は、まっ

たく同じ配列をもつはずである。しかし、同じバーコード配列をもつターゲット配列中にも、異なる配列をもつものがあつた(図)。

これはシーケンシング時の配列同定エラーが原因である可能性が高く、そのせいで、次世代シーケンサでの配列決定精度は、一般に、サンガーシーケンシングに比べてかなり低い。

		70	80
read4	AATCGGAGTTCTGCGCGATATCT		
read5	AATCGGAGTTCTGCGCGATATCT		
read3	AATCGGAGTTCTGCGCGATATCT		
read2	AATCGGAGTTCTGCGCGATATCT		
read6	AGTCGGAGTTCTGCGCGATATCT		
read1	AATCGGAGTTCTGCGCGATATCT		

ここでは、方法欄に記したように、多数決の原理を用いてエラーを同定し、そのエラーを除くことで、増幅前の配列を決定した。これにより、図の例では、左から2番目の配列は「A」となる（「G」がエラーと同定された）。

増幅されるものとして1塩基だけ異なる配列を用いて試したところ、このバーコードを利用してエラーを除く方法を用いて、1塩基の違いを区別できることが分かった。したがって、次世代シーケンサを用いた場合でも、バーコード法の利用により、1塩基の精度で配列を同定し、同時に1分子の分解能で分子数を定量できた。さらに、この方法を用いて、高い定量ダイナミックレンジを得ることもできた。

これらの実験において、さらに、分子バーコード法の新しい機能を示すことができた。通常、異なるサンプルを混ぜてシーケンシングする際には、サンプル毎にインデックスという配列を付加する。シーケンシング後、このインデックスの配列毎に同定されたターゲット配列をより分けて、どの配列がそのサンプル由来かを判断する。今回、バーコード配列を利用することで、インデックスの部分のシーケンシングエラーも同定し、エラ

一によるサンプルの混在を避けられる可能性を示すことができた。この結果は、分子バーコードの利用効果をより高めるものと考えられる。今後、さらに詳細を解析していく予定である。

ここれらの測定では、一度にたくさんの配列について解析が必要となるため、解析を自動化するソフトウェアを作製し、前後の解析と連続的に操作できる予備的なパイプラインも構築した。

よりエラーの影響を少なくするために、リニア増幅をしてから指数関数的に増幅する方法も試した。モデルサンプルを用いての増幅はできるようになったが、上記のエラー同定法でも十分精度が高いと考えている。

1分子での定量精度と1塩基での配列同定精度を持つ測定原理に加えて、解析パイプライン（現在は予備的なもの）も付随させた本計測システムは、核酸を対象にした様々なハイスループット計測で利用されていくと考えられる。今後は、解析パイプラインを最適化して実装し、測定システム全体として世界に広めていきたいと考えている。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）〔雑誌論文〕（計3件）

① 城口克之

デジタル RNA シークエンシング ゲノムワイドな1分子定量法

生物と化学（日本農芸化学会）、査読有、Vol. 54 No. 11 p787-788 (2016)

<https://katosei.jsbba.or.jp/index.php?id=678&bt=on>

② 城口克之、小川泰策

次世代シーケンサーを用いた DNA, RNA の網羅的デジタル計測、現代化学、査読無、2016年9月号 (No. 546) p47-50

<http://www.tkd-pbl.com/book/b244284.htm>

1

〔学会発表〕（計4件）

① 城口克之、Jianshi Jin

1細胞レベル・塩基分解能でのハイスループット菌叢解析、第11回日本ゲノム微生物学会年会、平成29年3月2日 慶應義塾大学（神奈川県、藤沢市）

② 城口克之

Finding genomic minority cells by sequencing、第54回日本生物物理学会年会、平成28年11月25日 つくば国際会議場（茨城県、つくば市）

③ 城口克之

High Throughput Single Cell Analysis for Target Genomic Sequences、International conference on Single Cell Research 2016”、平成28年11月16日、東京大学（東京都、文京区）

④ 城口克之

A Novel Method for High Throughput Single Cell Identification and Counting by Target Gene Sequencing、QBiC Symposium 2016”Decoding Organisms by Quantitative Cell Profiling”、平成28年9月6日、千里ライフサイエンスセンター（大阪府、豊中市）

〔その他〕

アウトリーチ活動

（1）理化学研究所生命システム研究センターの般公開にてサイエンスカフェに参加した。

（2）理化学研究所生命システム研究センターのスプリングコースにて、研究の紹介を行った。また、実習コースとして学生の受け入れを行った。

6. 研究組織

(1) 研究代表者

城口 克之 (SHIROGUCHI, Katsuyuki)

理化学研究所・生命システム研究センター・

ユニットリーダー

研究者番号：00454059