

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 29 日現在

機関番号：82401

研究種目：挑戦的萌芽研究

研究期間：2014～2014

課題番号：26640119

研究課題名(和文)ゲノムワイドRNA高精度絶対計数法の標準化

研究課題名(英文)Standardization of genome-wide absolute and accurate counting of RNA molecules

研究代表者

城口 克之(Shiroguchi, Katsuyuki)

独立行政法人理化学研究所・統合生命医科学研究センター・上級研究員

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)：申請者らは、次世代シーケンサを用いて、核酸の数をゲノムワイドかつ高精度でデジタル定量できる分子バーコーディング法を以前に開発した。本研究では、この方法を発展させ、バーコード配列のデザインなどを工夫することにより、低コストで高いダイナミックレンジを得ることに成功した。この成果は、本方法の世界標準化を加速すると考えられる。

研究成果の概要(英文)：We have previously developed the molecular barcoding method which enables one to perform accurate digital counting of the copy number of nucleic acid molecules genome-wide using next generation sequencer. This time, we developed it further, for example, by optimizing the sequence of molecular barcode, and succeeded in obtaining high dynamic range with less cost. These results may push to make our molecular barcoding method world-standard.

研究分野：生物物理学 定量オミクス

キーワード：高精度定量 デジタル定量 バーコーディング ゲノムワイド シーケンサ

1. 研究開始当初の背景

次世代シーケンサを用いた RNA シーケンシングが、遺伝子発現解析の重要なツールとなっている。同時に、これまで定性的な解析を目的としていたシーケンサを定量的な解析装置として用いる点で、定量精度における課題が生じている。

本研究者は、DNA バーコーディング法(後述)により、一分子の分解能でゲノムワイドに cDNA を絶対定量する方法を開発した(Shiroguchi et al. PNAS 2012)。この方法は、ライブラリー調整を工夫し、シーケンサは既存のものをそのまま使用するという簡易性も注目されている。

この技術を標準法として発展させ、広くサイエンスや医学の発展に寄与するためには、多くの研究者や臨床の現場などでも容易に利用できるようにすることが重要である。そこで課題となるのが、定量計測できる分子の最大数を増やすとコストが増加する点である(詳細は後述)。この課題が解決されれば、バーコーディング法の普及により、極めて精度の高い計測をベースとするサイエンスの展開や新しい診断法の開発などが期待される。例えば、コピー数が少ないと考えられている、分化に重要な転写因子の計測が可能となり、分化誘導法の評価などに貢献できると考えられる。

2. 研究の目的

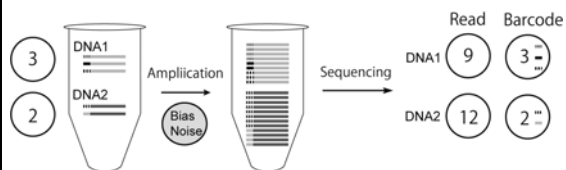
本研究では、分子バーコーディング法の初期コストが高いという難点を克服するため、革新的な技術を組み込み、現在 RNA シーケンシングを行っている方々が、この新しい方法を容易にかつ低コストで利用できる方法を開発する。新しい方法は、世界に広まる標準法になりうると考えている。さらには、本方法は、バ

ーコードを結合させることで、蛋白質分子などの高感度絶対定量計測にも応用でき、その発展性は極めて高いと考えている。

3. 研究の方法

(1) 本研究者らが開発した分子バーコーディング法の原理

現在広く行われている RNA シーケンシングでは、RNA を cDNA に変換し、PCR 増幅した cDNA の個数を次世代シーケンサで数えている。この時、PCR 増幅の過程で大きなノイズが生まれる(例えば1サイクル後に増幅されたものとされないものを比べると、その時点で2倍異なり、このノイズが複数サイクルにより増幅される)。本研究者らはバーコード(DNA 配列)を用いてこのノイズを除去した(図)。



図：バーコード配列と次世代シーケンサを用いた網羅的核酸デジタル定量法の原理：

左のチューブに2種類の配列をもつ DNA1 と DNA2 が存在し、それぞれコピー数が2、3だとする。各 DNA にそれぞれ異なるバーコード DNA を結合させ(各 DNA の左側の部分)、PCR で増幅し、バーコードと DNA の配列を次世代シーケンサで読む。PCR 増幅時に、配列に依存する増幅効率の差(バイアス)や増幅効率が100%ではないことに起因するノイズにより、増幅後の DNA の数の比は増幅前と異なる(図では DNA1:DNA2 = 9:12 = 3:4)。しかし増幅後もバーコードの数は DNA1 が3つ、DNA2 が2つであり、バーコード法により増幅後の測定から増幅前の DNA の絶対数を高精度で決定できる。

この方法で正確に定量するためには、それぞれの DNA が異なるバーコード配列を結合する必要がある。したがって、バーコードの種類数が、定量できる DNA の個数の上限を決める。本研究では、この上限を低コストで高めることにチャレンジした。

(2) 研究の流れ

バーコードとしてランダム配列を用い、“一種類”のバーコード配列を購入する費用で足りるようにした。実際には“NNNNNN....”という配列を発注するが、コスト的には一種類の配列として扱われ、すべての種類の配列を含む DNA が一本のチューブに入れられて届く (“N”は A or G or C or T を意味する)。バーコード配列を長くして極めて多くのバーコード配列を準備し、実際に使われたバーコード配列同士が、(統計的に) より似ていない配列になるようにした。これにより、生じたエラーがどのバーコード配列から起きたかを同定した。具体的には、エラー配列を含めてクラスター化し、クラスターの数を増幅前の DNA の数とした。この方法で決定した DNA の分子数が、増幅前の分子数に一致するかを確認した。

(3) バーコード設計による技術的な問題点

バーコードを用いたデジタル計測で大きな問題となるのは、ポリメラーゼによる複製反応やシーケンシングで起こるエラーである。例えばあるターゲット分子に AAAAAA というバーコードが結合し、それがエラーを起こして TAAAAA という配列を結果として得ると、もともと二つの分子が存在したと解釈する。本研究者が以前に報告した方法ではバーコード配列を事前に選択した。例えば、

AAAAAA という配列はバーコードとして用い、TAAAAA という配列は用いていないことが既知である。したがって、TAAAAA はエラーだと同定でき、分子数としてカウントしないため、エラーが結果に影響を与えない。

本研究では、ランダムバーコード(前述した“NNNNNN....”)を用いた。6塩基を例にすると、 4^6 種類の配列が一つのチューブ内に存在する。バーコードを選択する場合には一つ一つの配列を指定して多種類の配列を購入するが、ランダムバーコードを利用すれば、先述したように NNNNNN.... という“一つ”の配列を購入すればよく、劇的な低コスト化が実現する。ここで問題となるエラー問題を下記の方針で解決した。

(4) エラー問題の解決方針

ランダムバーコードのエラー問題を解決するため、長いバーコード配列と、クラスター解析法を用いた。例えば、NNNNN.... と 20塩基つながる配列をバーコードとする。このとき、 10^{12} (~ 4^{20})種類のバーコード配列が存在している。チューブ内に 10分子の DNA が存在し、これらにこのバーコード配列を結合させたとする。すると、結合したバーコード配列は、統計的に、同じ、もしくは似ている確率は非常に小さい。例えば、結合した一つのバーコード配列が、AAAAAAA.... とすべて A だったとする。この時、増幅してシーケンシングをすると、TAAAAAAA.... というエラーも起こりうるし、AGAAAAAAA.... というエラーも起こり得る。しかし、これらの配列が、他に使われたバーコード配列やそのエラー配列と一致する可能性は極めて低い。したがって、エラーを含めてクラスター化すると、AAAAAAA....、TAAAAAAA、

AGAAAAA...などの似ている配列は一つのクラスターを形成する。これにより、クラスターの数をもともと存在した分子数に一致する。この方法を実験的に、そして定量的に評価した。

具体的には、ランダムバーコードを設計し、固定配列とランダムバーコードがデザインされたオリゴ DNA を複数購入した。これらの複数のオリゴ DNA を、既知の数、1つのチューブに入れ、PCRで増幅した。そして、増幅産物の配列を次世代シーケンサを用いて解読した。その後、バーコード配列をクラスター化し、クラスター数をもとめた。

4. 研究成果

クラスター数は、増幅前にチューブに加えた分子数とよく一致した。この結果により、適切にバーコードをデザインし、適切な解析を行うことで、ランダムバーコードを用いて核酸の数をデジタル定量できることが示された。低コストで高いダイナミックレンジを得ることができたという成果を世界に発信することで、この技術の標準化が加速されると考えている。

さらに、現在注目されている一細胞シーケンシングにおいても、このデジタル計測法を用いれば、劇的な定量精度の向上が期待できる。

招待講演を含めて複数の学会で成果を発表した。現在、論文投稿の準備をしている。

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 件)

[学会発表] (計 7件)

- ① Katsuyuki Shiroguchi, Genome-wide Digital mRNA Counting Using "Molecular Barcoding" and Its Application for Differentiation Study, BIT' s 5th

Annual World Congress of Molecular & Cell Biology-2015、平成 27 年 4 月 26 日、Nanjing (China)

- ② 城口克之、100 cell transcriptome analysis: highly accurate absolute digital quantification with single molecule resolution using molecular barcoding、第 43 回日本免疫学会学術集会、平成 26 年 12 月 12 日、国立京都国際会館 (京都府・京都市)

- ③ 城口克之、Towards next generation system-wide measurements: Highly accurate and absolute quantification of RNA molecules from small amounts of sample by "molecular barcoding"、第 37 回日本分子生物学会年会、平成 26 年 11 月 26 日、パシフィコ横浜 (神奈川県・横浜市)

- ④ Katsuyuki Shiroguchi、Accurate system-wide gene expression analysis with single molecule resolution by using molecular barcoding、MicroRNAs & Single Molecule Biology/Genome Editing Europe Symposia-2014、平成 26 年 11 月 3 日、Cambridge (UK)

- ⑤ 城口克之、Absolute genome-wide quantification of gene expression with single molecule resolution using molecular barcoding、第 52 回日本生物物理学会年会、平成 26 年 9 月 27 日、札幌コンベンションセンター (北海道・札幌市)

- ⑥ Katsuyuki Shiroguchi、Highly Reproducible and Sensitive Genome-Wide Digital RNA Counting using Molecular Barcoding、Gordon Research Conferences Rare Cells in Circulation、平成 26 年 8 月 3 日～平成 26 年 8 月 7 日、South Hadley (USA)

- ⑦ 城口克之、RNA 分子数の絶対定量により状態変化を捉える：分子バーコーディング法による RNA ゲノムワイドデジタル計測法の開発とその応用、第 66 回日本細胞生物学会大会、平成 26 年 6 月 11 日、奈良県新公会堂 (奈良県・奈良市)

[図書] (計 件)

[産業財産権]

○出願状況 (計 件)

名称：

発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

城口 克之 (SHIROGUCHI, Katsuyuki)
理化学研究所・統合生命医科学研究センター・上級研究員
研究者番号：00454059

研究者番号：

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：