

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年6月11日現在

機関番号：12601

研究種目：若手研究（B）

研究期間：2010～2011

課題番号：22710182

研究課題名（和文）出芽酵母の転写単位確定とノンコーディングRNA発現の全貌把握

研究課題名（英文）High-throughput definition of transcriptional units to grasp global expression states of non-coding RNA in budding yeast cells

研究代表者

三浦 史仁 (MIURA FUMIHITO)

東京大学・大学院理学系研究科・特任助教

研究者番号：50447348

研究成果の概要（和文）：出芽酵母において mRNA の 5' 末端と 3' 末端を同時に確定することを試みた。対数増殖期の細胞と減数分裂誘導直前と誘導後 2 時間毎に 6 回の回収を行った細胞を用意した。それぞれから独自に開発したクローニングフリー GIS 法により配列決定のための鋳型を調製し、次世代シーケンサで配列決定を行った。このデータを用いて、出芽酵母遺伝子の転写開始点と終結点の双方の情報を同時に格納したユニークなデータベースを作成している。

研究成果の概要（英文）：Simultaneous determinations of 5' and 3' terminal sequences of a mRNA molecule were conducted on budding yeast *Saccharomyces cerevisiae* with highly parallelized manner. Logarithmically growing cells, and cells before and six times every 2 hours after induction of sporulation were collected. Cloning-free GIS were applied for then, and sequencing was performed on a next-generation sequencer. Construction of a unique database storing transcription initiation and termination position of each budding yeast transcript as a pair is currently underway.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	2,900,000	1,170,000	4,070,000
2011年度	1,000,000	300,000	1,300,000
年度			
年度			
年度			
総計	3,900,000	1,470,000	5,370,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・ゲノム生物学

キーワード：トランスクリプトーム

1. 研究開始当初の背景

真核生物の細胞内では多くのゲノム領域で転写が起こっており、これらの転写物にはタンパク質へと翻訳されることのないノンコーディング RNA (ncRNA) が多く含まれていると考えられている。ncRNA の発現が転写機構のノイズによるものか、細胞機能にとって意味のある現象なのかは現在も議論

が続いているが、一部の ncRNA は mRNA の翻訳制御やヒストンの化学修飾変化に関わるなど細胞機能に重要な役割を果たしていることが知られている。

モデル生物である出芽酵母 *S. cerevisiae* では、高等真核生物における ncRNA の最も代表的な機能の一つである RNAi の機構こそないものの、転写干渉による転写制御、転写

と共役したヒストン化学修飾など転写そのものに対してシスに働く ncRNA の機能の存在が数年前には明らかにされていた。最近になって、自身が転写された領域以外のゲノム領域の発現制御に関わるいわゆるトランスな ncRNA の発現制御機構の存在が示唆され、また近縁種 *S. castellii* や *C. albicans* では RNAi の機構が同定されている。これらの知見は、ncRNA の発現と機能が真核生物に普遍的に成り立つ現象であることを意味しており、ncRNA の普遍的な機構の解析を、より扱いが容易なモデル生物である *S. cerevisiae* やその近縁種を用いて効率的に行うことが可能であることを示唆していた。

出芽酵母の細胞内にどれだけの種類の ncRNA が存在するのかは、高密度タイピングアレイを用いた転写物の解析や我々が行った完全長 cDNA ライブラリの 5' 側ワンパスシーケンス解析、あるいは RNA-Seq 解析のデータからある程度の推測をすることができる。これらの解析では、出芽酵母ゲノム中の遺伝子間領域に 1000 種を超える新規転写物が同定され、そのうち 1 割程度は ORF が組めない完全な ncRNA であることが分かった。また、およそ 2000 個の遺伝子座にアンチセンス転写物が同定されている。センス鎖上から転写される ncRNA の存在もあわせて考えると、6000 遺伝子を有する出芽酵母の細胞内にはそれと同程度の ncRNA が発現している可能性も考えられる。しかし、出芽酵母細胞内における ncRNA の転写の意義やその機能を探る足掛かりを得るためには、確たる手法を用いた実測によって ncRNA 発現の全貌を明らかにする必要がある。

ncRNA の存在を示すためにはそこに転写物が存在することを示すだけでは不十分で、当該転写物が近隣遺伝子からは独立したひとつの転写単位であることを示さなければならない。また、比較的発現量の低いことが多い ncRNA を効率的に検出するためには、高感度な解析技術の導入が必要不可欠である。我々は先に行った出芽酵母の完全長 cDNA ライブラリの解析において、5' 側からのワンパスシーケンシングのみでは 3' 末端側の情報が不足しているため、配列決定されたクローンの既知遺伝子への帰属の可否の判定が困難である例を多く経験した。また、大腸菌へのクローニングを介した cDNA ライブラリの解析は多くの利点があるものの、コスト面や配列決定の深度に実用上の限界があった。そこで転写物の 5' と 3' の末端タグ配列を同時に取得することにより転写単位を確定することが可能な Gene identification signature (GIS) 法に注目し、サンプリングのバイアスとなる 3 度の大腸菌へのクローニングステップを必要とする原法の問題点を克服した Cloning-free GIS 法

を開発してきた。Cloning-free GIS では、全ての操作を試験管内で行うことによりクローニングによるサンプリングバイアスを回避しているほか、水/油エマルジョン中で cDNA の環状化を行うことにより、サイズ依存的な環状化のバイアスを低減しており、様々なサイズの DNA 断片が混在する cDNA ライブラリを効率的にペアエンドダイタグ (Paired-end ditag) 化することが可能である。Cloning-free GIS 法と次世代シーケンサの能力を組み合わせることにより、より多くの転写物の転写範囲を確定することが可能となり、より高感度に ncRNA の転写単位を判別することが可能になると考えられる。これまで我々は出芽酵母の約 5000 遺伝子の発現量を正確に絶対定量し、出芽酵母の細胞内には培養条件によって 1 万 5 千コピーから 3 万コピー程度の mRNA が存在することを明らかにしているが、研究開始当時の次世代シーケンサの解読能力は 1 ラン当たり 1 億リードであるため、1 回の配列決定操作で 1000 細胞に 1 コピー存在する極めて低い発現量の転写物が検出可能であると考えられた。

2. 研究の目的

本研究では、研究代表者自身が開発を行ってきた Cloning-free GIS 法と次世代シーケンサを組み合わせ、出芽酵母細胞内で発現している転写物を網羅的に解析し、各転写物の転写範囲の確定を行うことを第 1 の目標とした。得られた転写物の転写範囲情報をデータベース化し、ゲノムブラウザを通じて可視化した上でインターネットを通じて公開することを第 2 の目標とした。転写物のなかで特に ncRNA に注目し、他のゲノム上にマップされたオーミクスデータ、特にヌクレオソーム構造やヒストンの化学修飾状態との関連性を調査することを第 3 の目標とした。

3. 研究の方法

対数増殖期および減数分裂期の出芽酵母細胞を回収し、それぞれから抽出した RNA を元に Cloning-free GIS 法によって完全長 cDNA の 5' と 3' の配列のペアエンドダイタグライブラリを調製することを試みた。得られたライブラリの一部をライフテクノロジー社 SOLiD システムあるいはイルミナ社 MiSeq システムにより読み出した。PC クラスタシステムを駆使して得られた配列の参照配列へのマッピングを試みた。5' 側および 3' 側配列を決定することにより各転写物の転写範囲を確定し、データベースへと格納する試みは継続中である。今後、各転写物の転写範囲情報に基づき、それぞれを既知遺伝子へ帰属し、既知遺伝子の転写範囲情報を確定する予定である。帰属が困難な転写物に関しては、その転写物が翻訳される可能性を配列解

析により明らかにすると同時に ncRNA の候補としてリストアップし、ヌクレオソーム位置やヒストンの化学修飾状態などの他のオミクスデータとの比較を行う必要があるだろう。そこで回収された細胞から MNase-Seq ライブラリを調製し、上記 GIS ライブラリと同様に SOLiD システムあるいは MiSeq システムにより配列決定を行った。現在これらのデータの参照ゲノム配列へのマッピングを行っている。

4. 研究成果

(1) 本研究の開始当初、出芽酵母の対数増殖期の細胞と減数分裂期の細胞は比較的容易に得られるものと考えられていた。しかし、原因は後に細胞株の変性であることが明らかになるのだが、実際に培養を行ってみると減数分裂期の細胞を十分な再現性と回収量で得ることが出来なかった。そこで減数分裂誘導条件の再確認を行うことと並行して、より少ない開始材料から Cloning-free GIS ライブラリを調製可能にするための技術改良を行った。結局、細胞株をバンクから新たに取り寄せることで十分な減数分裂期の細胞量を確保することが可能になったが、この過程で Cloning-free GIS 法の必要開始量をそれまでの 20 μ g から 1 μ g へと減らすプロトコルの高感度化を実現することができた。得られたライブラリの配列決定には新学術領域「ゲノム支援」による支援により、ライフテクノロジー社の SOLiD システムを用いたが、機器の不具合により十分な質と量の配列情報を得ることが出来なかった。その後別経費で所属研究室に措置されたイルミナ社 MiSeq システムで配列決定を行うことが可能になったが、配列決定の際のアダプター導入にトランスポゾームを利用したタグメンテーション法を採用すると、より容易かつ高感度にアダプター配列を導入することが可能となり、プロトコルの安定化を実現することができた。また、それまではタイプ IIS 型制限酵素である MmeI によるタグ切り出しを採用していたため、20 塩基程度の短いタグ配列しか読み出すことが出来なかったが、トランスポゾームを利用する系を用いた結果、より長いタグ配列を回収することが可能になり、得られる情報量が向上した点は技術面での大きな成果となった。

(2) これまでに出芽酵母の対数増殖期の細胞と減数分裂誘導前から誘導後 2 時間間隔で 12 時間まで経過した細胞を回収し、クローニングフリー-GIS によるライブラリ調製を行った。同時に同じ細胞から MNase-Seq ライブラリも調製を行った。これらのうちの一部のライブラリはイルミナ社 MiSeq システムによる配列決定に供されている。Cloning-free GIS のライブラリから得られた粗データには転

写物の 5' 末端と 3' 末端配列のほか cDNA を環状化する際に必要なアダプター配列等が含まれているが、これらの配列中から転写物の配列情報を抜き出すプログラムの開発を行った。抜き出された配列をゲノム配列上にマッピングする操作を現在継続して進めている。一部のタグペアを個別に解析してみると、確かに既存遺伝子の 5' 端と 3' 端をとらえられていることが確認できたことから、調製されたライブラリは転写物の範囲決定に有効な情報をもたらすものと考えられる。MNase-Seq ライブラリは同じく MiSeq システムによるペアエンドシーケンシングで配列決定を行った。この結果得られた配列は既存ツールの SOAPAligner により効率的にゲノム配列上にマッピングされていた。今後、配列決定を終えていない残りのライブラリの配列を決定するとともに、上記解析系の改良を行い、ゲノムブラウザでの可視化等を行う必要がある (図 1)。

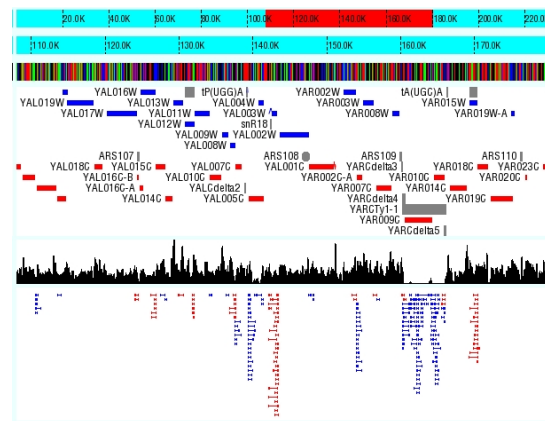


図 1、ゲノムブラウザによる可視化 Cloning-free GIS と MNase-Seq のデータを可視化した例。各トラックは上から染色体上の表示位置を示す Overview Track、表示位置のスケールを表す Ruler Track、塩基組成を表す Base Color Track、遺伝子位置を表す Feature Track、MNase-Seq のタグの張り付き具合を示す Graph Track、Cloning-free GIS のダイタグの張り付き具合を示す Ditag Track である

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 2 件)

- ① 三浦史仁、伊藤隆司、進化したメチローム解析技術、医学の歩み、査読なし、236 巻 6 号、2010 年、643-651

- ② 三浦史仁、伊藤隆司、ゲノム網羅的メチル化解析、実験医学増刊「エピジェネティクスと疾患」、査読なし、28 巻 15 号、2010 年、2407-2414

〔学会発表〕(計 5 件)

- ① 三浦史仁、伊藤隆司、ゲノム網羅的バイサルファイトシーケンシングの高感度化への取り組み、第 34 回日本分子生物学会年会、2011 年 12 月 13-16 日、横浜
- ② 三浦史仁、榎本悠佑、伊藤隆司、サブナノグラムからはじめるゲノム網羅的バイサルファイトシーケンシング、日本エピジェネティクス研究会第 5 回年会、2011 年 5 月 19-20 日、熊本
- ③ Fumihito Miura, Yusuke Enomoto, Ryo Dairiki and Takashi Ito, Whole genome bisulfite sequencing from subnanogram quantities of DNA, The 9th International Workshop on Advanced Genomics, 2011 年 7 月 12-14 日、東京
- ④ 三浦史仁、榎本悠佑、伊藤隆司、サブナノグラムからはじめるゲノム網羅的バイサルファイトシーケンシング、第 33 回日本分子生物学会年会、2010 年 12 月 7-10 日、神戸
- ⑤ 三浦史仁、榎本悠佑、伊藤隆司、サブナノグラムからはじめるゲノム網羅的バイサルファイトシーケンシング、日本エピジェネティクス研究会第 4 回年会、2010 年 5 月 28-29 日、米子

〔その他〕

ホームページ等

<http://itolab.cb.k.u-tokyo.ac.jp/cgi-bin/DesktopTracks/bin/Browser>

6. 研究組織

(1) 研究代表者

三浦 史仁 (MIURA FUMIHITO)
東京大学・大学院理学系研究科・特任助教
研究者番号：50447348

(2) 研究分担者

なし

(3) 連携研究者

なし