

完全長 cDNA ライブラリーを利用したトランスクリプトーム解析と技術開発

●菅野 純夫¹⁾ ◆中井 謙太²⁾ ◆橋本 真一³⁾

1) 東京大学新領域創成科学研究科 2) 東京大学医科学研究所 3) 東京大学医学系研究科

<研究の目的と進め方>

遺伝子発現ネットワーク研究の基盤の形成を目指し、特定領域研究「生命システム情報」及び特定領域研究「比較ゲノム」と連携し、完全長 cDNA ライブラリー、5' 端濃縮ライブラリー、5' 端 SAGE(5' SAGE) ライブラリーなどを利用したトランスクリプトーム解析を行い、遺伝子の転写開始点の詳細とそのコアプロモータ領域のデータベース化をおこない、さらに、これらトランスクリプトーム解析に関連した技術開発を進める。

1, 特定領域研究「生命システム情報」及び特定領域研究「比較ゲノム」と連携し、ゲノム研究関連特定研究領域で解析の対象となる真核生物であるメダカ(近縁種2種)、ホヤ(各種)、ナメクジウオ、ギボシムシ、霊長類(各種)、カイコ、コムギ、クラミドモナス等を中心に、オリゴキャップ法による完全長 cDNA ライブラリー、5' 端濃縮ライブラリー、5' 端 SAGE(5' SAGE) ライブラリーを作製し、それらの cDNA ライブラリーからのクローンの5' 端 EST 配列を決定することで、遺伝子の転写開始点の詳細とそのコアプロモータ領域を同定する。

2, コアプロモータ領域の配列情報と EST、SAGE、マイクロアレイ等の遺伝子発現頻度情報を組み合わせた、コアプロモータ領域と転写制御の比較ゲノム解析を行うためのデータベースの作製をおこなう。

3, 上記の課題に関連し、技術開発を進める。

<2008 年度の研究の当初計画>

1, 特定領域研究「生命システム情報」及び「比較ゲノム」の班員より、RNAの提供を受け、完全長cDNAライブラリー、5'端濃縮ライブラリーを作製し、5'端ESTの決定を行う。大規模EST配列決定は大量DNAシーケンス支援班と連携しておこなう。解析対象の選択については、他3領域との密接な連携のもとに、適切かつ柔軟に対応する。

2, 技術開発において、まず、現在既に開発に成功している5'-end Solexa法の確立と、使用RNAの微量化とに注力する(橋本)。また、5'-end Solexa法を利用したnon-coding RNAの同定及び定量化の開発を目指す。さらに、新型シーケンサーを使用したcDNA全長配列決定法の開発を目指す(菅野、中井)。大量の5'-end Solexa法用ライブラリー作製のために、研究補助員またはポストドクを1名必要とする。

3, 5'-end Solexa法などにより大量に生産される5'tagデータを、現在作製運用中のDBTSSに取り込み、既存の遺伝子発現頻度情報とコアプロモータ領域の配列情報を組み合わせた、比較ゲノム解析のデータベースの枠組みと、そのために必要なツール群を作る(中井)。データベースやソフトウェアは情報解析支援班と連携して作成する。

4, 5'-end Solexa法で得られる多量のデータを、RNA同定とプロモーターマッピングのみならず、定量的発現解析に利用できる

いか、全ゲノム配列がしっかりと決定したヒトを材料に検討する。

これらの研究で得られた成果は、国内のみならず、コールドスプリングハーバーのゲノム会議やHUGO、HUGOpacificの総会等の国際会議で発表を行う。

<2008 年度の成果>

現在までに、「比較ゲノム」と連携し、ゼニゴケ福澤秀哉(京大)、タテエリベンモウチュウ岩部直之(京大)、ショウジョウハエ垣敏郎(首都大)、アピコンプレクサ原虫 渡辺純一(東大)、メダカ堀寛(名大)等につき、総計で、完全長 cDNA ライブラリー15種、5' SAGE ライブラリー8種類を作製した。一部のライブラリーについては、5' 端の配列決定を進め、約10万のデータを得ている。さらに、それ以上のEST配列決定では、小原グループと連携している。

また、次世代シーケンサーを用いたトランスクリプトーム解析について、いくつかの方法論の開発を行った。1)オリゴキャップしたcDNAを次世代シーケンサーを用いて配列決定することにより、数日のうちに数千万の転写開始点を決定する方法を開発した。2)クロマチン免疫沈降したDNAサンプルを次世代シーケンスで配列決定するChip-Seq法のプロトコルを確立した。3)1000クローン程度のクローンDNAを混ぜてショットガンシーケンスを行うことにより、安価迅速に全長配列を決定する方法を開発した。

このように開発した方法を用いて、1)現在までに、種々のヒト培養細胞の5'tagを、合計で、約4億tag得ている。特に、特定の遺伝子に属するtag数が遺伝子発現レベルと相関することを確認し本法が遺伝子発現解析ツールとしても使用できることを示した(下記成果発表の1参照)。2)約2万のcDNAクローンの配列決定を行った。

また、1)について、約1億tag得られた各の転写開始点データについて、必要なプラットフォームの整備を行ったうえ、DBTSSを通じて公開した。情報解析では、新型シーケンサーによる5'SAGEタグ解析システムを作成し、メダカゲノムプロジェクトに利用した。

<国内外での成果の位置づけ>

完全長 cDNA ライブラリー、5'SAGE ライブラリーでは、本研究グループは、国内外で優位に立っている。われわれのグループと同等の能力を持つのは、理研の林崎グループだが、彼らはゲノムネットワークプロジェクトでヒトに注力しており、ヒト・マウス以外の生物は、われわれのグループがカバーしている。外国の研究グループからの共同研究の申し出も多く受け、可能な範囲で協力を行っている。

<達成できなかったこと、予想外の困難、その理由>

特になし

<今後の課題>

引き続き、次世代シーケンサーの利用法を開発していく必要がある

現在、試みているのは、

- 1) DNA メチル化の解析法
- 2) クロマチンの DNA 分解酵素感受性部位の網羅的解析
- 3) 短鎖 RNA や polyA の無い RNA の解析
- 4) ゲノムの一部の濃縮法
- 5) 短い塩基配列を使った新規配列の決定

などである。

この中で、2は、ほぼ問題なく出来つつある。4は、サンプルの調整の難しいものであるが、評価も難しい。5は、情報解析での課題であるが、ヒトでは難しい。1や3は、両方で課題がある。

次世代シーケンサーを用いた 5'SAGE 法はわれわれのグループに世界的に強みがあるので、情報解析のパイプラインも含め、総合的なプラットフォームを構築出来ればと考えている。しかし、大量のデータを処理し始めると、ヒト標準配列の不備(?)、シーケンサーのエラー率の問題等で、自身で解決できない部分が出てきている。世界的な連携も必要であろう。

また、開発した方法の使用法も含め、「比較ゲノム」と連携し、ヒト以外の生物でも他の追随を許さない情報を得るようにしていきたい。

<成果公表リスト>

1. 0901111023 (論文)

Hashimoto S, Qu W, Ahsan B, Ogoshi K, Sasaki A, Nakatani Y, Lee Y, Ogawa M, Ametani A, Suzuki Y, Sugano S, Lee CC, Nutter RC, Morishita S, Matsushima K. High-resolution analysis of the 5'-end transcriptome using a next generation DNA sequencer. *PLoS ONE*. 2009;4(1):e4108. Epub 2009 Jan 1.

2. 0901111011 (論文)

Wakaguri H, Suzuki Y, Katayama T, Kawashima S, Kibukawa E, Hiranuka K, Sasaki M, Sugano S, Watanabe J. Full-Malaria/Parasites and Full-Arthropods: databases of full-length cDNAs of parasites and arthropods, update 2009. *Nucleic Acids Res*. 37:D520-D525, 2009 Nov 4.2008 [Epub ahead of print]

3. 0901111005 (論文)

Yoshikawa F, Sato Y, Tohyama K, Akagi T, Hashikawa T, Nagakura-Takagi Y, Sekine Y, Morita N, Baba H, Suzuki Y, Sugano S, Sato A, Furuichi T. Opalin, a transmembrane sialylglycoprotein located in the CNS myelin paranodal loop membrane. *J Biol Chem*. 283:20830-20840, 2008

4. 0901110953 (論文)

Lee YS, Choi SL, Kim TH, Lee JA, Kim HK, Kim H, Jang DJ, Lee JJ, Lee S, Sin GS, Kim CB, Suzuki Y, Sugano S, Kubo T, Moroz LL, Kandel ER, Bhak J, Kaang BK. Transcriptome analysis and identification of regulators for long-term plasticity in *Aplysia kurodai*. *Proc Natl Acad Sci U S A*. 105:18602-18607 2008

5. 0901110945 (論文)

Takeda J, Suzuki Y, Sakate R, Sato Y, Seki M, Irie T, Takeuchi N, Ueda T, Nakao M, Sugano S, Gojobori T, Imanishi

T. Low conservation and species-specific evolution of alternative splicing in humans and mice: comparative genomics analysis using well-annotated full-length cDNAs. *Nucleic Acids Res*. 36: 6386-6395, 2008.

6. 0806171343 (論文)

Uno Y, Suzuki Y, Wakaguri H, Sakamoto Y, Sano H, Osada N, Hashimoto K, Sugano S, Inoue I. Expressed sequence tags from cynomolgus monkey (*Macaca fascicularis*) liver: A systematic identification of drug-metabolizing enzymes. *FEBS Lett*. 582:351-358, 2008..

7. 0806171340 (論文)

Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, Inoue I, Hida M, Suzuki Y, Sugano S, Terao K, Kusuda J, Takahashi I. Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. *BMC Genomics*. 2008 9:90

8. 0806171337 (論文)

Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 24:167-177, 2008..

9. 0901111151 (論文)

Oyama M, Kozuka-Hata H, Tasaki S, Semba K, Hattori S, Sugano S, Inoue JI, Yamamoto T. Temporal perturbation of tyrosine-phosphoproteome dynamics reveals the system-wide regulatory networks. *Mol Cell Proteomics*. 2008 Sep 24. [Epub ahead of print].

10. 0806171335 (論文)

Yamashita R, Suzuki Y, Takeuchi N, Wakaguri H, Ueda T, Sugano S, Nakai K. Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. *Nucleic Acids Res*. 36: 3707-3715, 2008.

2) データベース公開 (DB名、URL、簡潔な内容紹介)

DB名: DBTSS

URL: <http://dbtss.hgc.jp/>

各種生物の転写開始点の情報をデータベースとしたもの。現在は、ヒトとマウスが中心。なお、次世代シーケンサーを利用した大量 5'tag データも download 可能にしている。