

平成22年06月07日現在

研究種目: 特定領域研究

研究期間: 2005～2009

課題番号: 17020002

研究課題名(和文) 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

研究課題名(英文) Transcriptome analysis using full-length enriched cDNA library and technology development

研究代表者

菅野 純夫(SUGANO SUMIO)

東京大学・大学院新領域創成科学研究科・教授

研究者番号: 60162848

研究成果の概要(和文): われわれが開発したオリゴキャッピング法による完全長 cDNA ライブラリーを基盤に、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、多種類の生物の完全長 cDNA リソースの整備とトランスクリプトーム解析を行った。同時に、次世代シーケンサーとオリゴキャップ法を組み合わせ、ゲノムワイドに転写開始点を同定し、その発現量を半定量的に測定する方法を確立した。

研究成果の概要(英文): Using full-length cDNA library made by our Oligo-capping method as the technological base, we made large cDNA resource as well as transcriptome map of various organism with the collaboration with the Priority area of “Systems Genomics” and “Comparative Genomics”. At the same time, we developed a method for genome wide identification of the transcriptional start sites and for measurement of it strength in semi-quantitative manner, by combining Oligo-capping method with second generation sequencers.

交付決定額

(金額単位: 円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-------------|------|-------------|
| 2005年度 | 123,200,000 | 0 | 123,200,000 |
| 2006年度 | 123,700,000 | 0 | 123,700,000 |
| 2007年度 | 124,200,000 | 0 | 124,200,000 |
| 2008年度 | 125,600,000 | 0 | 125,600,000 |
| 2009年度 | 121,200,000 | 0 | 121,200,000 |
| 総計 | 617,900,000 | 0 | 617,900,000 |

研究分野: 複合新領域

科研費の分科・細目: ゲノム科学・応用ゲノム科学

キーワード: ゲノム 発現制御 バイオテクノロジー 蛋白質 生体分子

1. 研究開始当初の背景

われわれが開発したオリゴキャッピング法による完全長 cDNA ライブラリーより得られた完全長の cDNA 配列情報とゲノム配列情報と合わせることで、効率よくヒトの遺伝子のプロモータ領域を明らかにすることを旨とし、約 8000 の遺伝子プロモータ領域を明らかにし、DBTSS として公開し

た。それらと SNP 情報等の多型情報を組み合わせることで、プロモータ領域に変異を持つ疾患関連遺伝子を体系的に探査する上で有用な、独自の基盤を与えることが出来た。また、ヒトの完全長 cDNA ライブラリー及び 5' 端ライブラリーを多数作製し、その 5' 端 EST 配列を多数決定し、本プロジェクトのための技術基盤と設備的な基

盤ができあがった。

2. 研究の目的

- 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 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、以下の技術開発を進める。
 - (ア) 完全長 cDNA ライブラリーの改良。特に使用 RNA の微量化。
 - (イ) オリゴキャップ法を利用した 5' 端 SAGE ライブラリー作製法の開発。これにより、転写開始点同定の100倍程度高速化が予想される。
 - (ウ) 遺伝子発現解析用オリゴアレイの設計法の開発。ショウジョウバエや線虫、ヒトのようにゲノムが完全にわかっている場合と、ゼブラフィッシュのようにドラフトのみの場合に分けて、最適の設計アルゴリズムを検討し、実験で評価を行う。
 - (エ) non-codingRNA の効率的クローン化のための、新しい cDNA ライブラリー作製法の開発。
 - (オ) mRNA 量の絶対定量法、スプライシングパターン同定法の開発

3. 研究の方法

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。

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' 端 SAGE ライブラリー作製法の確立と、使用 RNA の微量化とに注力する。また、5' SAGE 法を利用した non-coding RNA の同定及び定量化の開発を目指す。

3, 現在作製運用中の DBTSS を中核に、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。

4, まず、全ゲノム配列が決定した例としてヒトを、次にドラフト配列のみである場合の例としてゼブラフィッシュをターゲットに、遺伝子発現解析用オリゴの最適の設計アルゴリズムを検討する。

5, 技術開発において、まず、現在既に開発に成功している 5' 端 SAGE ライブラリー作製法の確立と、使用 RNA の微量化とに注力する。また、新型超並列シークエンサーを使用した 5' 配列決定の高速化を実現するために、特別な 5' SAGE 法を開発する。

6, 新型超並列シークエンサーから得られる大量の短鎖配列をゲノム配列へのマッピングの最適アルゴリズムおよび最速アルゴリズムを開発検討し、実際のデータを使って評価を行う。

4. 研究成果

1, 現在までに、「比較ゲノム」と連携し、ヒメツリガネゴケ 長谷部光泰(基生研)、ゼニゴケとクラミドモナス 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 相垣敏郎(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17種、5' SAGE ライブラリー11種類を作製した。

さらに、完全長 cDNA ライブラリーをワラビー8種、カイコ1種、ヒトデ2種、馬4種、クラミドモナス4種、ゼニゴケ4種、ヒメツリガネゴケ1種、ゼブラフィッシュ4種、ウシ4種について作成した。また、5' SAGE library を、ヒメツリガネゴケ4種、ハエ1種について作成した。一部のライブラリーについては、5' 端の配列決定を進め、計30万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た。

2, メダカに関しては、最終的に、1,186,742個の5' SAGE tags を収集してメダカゲノム上に転写開始点を同定し、20,141個の遺伝子を予測した。うち3,727個は他生物種の遺伝子とホモロジーがなく、無作為に抽出した194遺伝子を検証した結果約2,900は新規遺伝子であることが推定された。

3, DNA チップ用のオリゴヌクレオチドプローブをヒト、マウス、出芽酵母の遺伝子について設計し、ヒトについては web からア

カデミアに公開した。その他は東大 TLO を通じて企業にライセンスし、たとえば DNA チップ研究所から製品化されている。

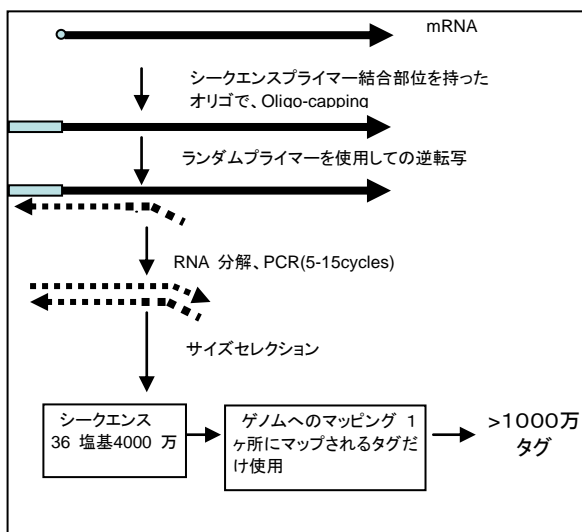
4、また、5' SAGEライブラリー作製の効率化し、コスト、時間などをそれぞれ半分以下にし、5' SAGE法に使用するRNAの微量化にも成功(5' 側の特異的RNA増幅法により、ngレベルのtotal RNAからライブラリーを製作する。)した。さらに、従来の5' SAGE法は転写開始点から20baseを特定することしか出来なかったが、25baseを特定出来る方法を開発した。

5、情報解析では、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。さらに設計が困難であったゲノム領域についてのマイクロアレイオリゴプローブの設計法の開発やMultiplex genomic PCRの設計方法を開発し、ウェブサーバーから公開した。

6、新しい超並列タイプの454シーケンサーによる高効率5' 端塩基配列決定の条件決定をした。新型シーケンサーでcDNAを読んだ例は世界的になくこれが初めて試みである。新型の超並列シーケンサーを用いた、安価で高速な転写開始点決定法を確立することは今後のゲノム研究にとって極めて重要であると考えている。

さらに、イルミナ社の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、いくつかの方法論的開発を行った。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。

図1 次世代シーケンサーを用いた5' 端の解析



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サーで配列決定する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シングを行うことにより、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。

2) 約8万の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を用い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。

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 44 件) すべて査読有り。

1)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:D520-525, (2009) (901111011)

2)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, 4:e4108, (2009) (901111023)

3)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 USA, 105:18602-18607, (2008) (901110953)

4)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) (901110945)

5)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, 008 Feb 24;9:90 (806171340)

6)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) (806171337)

7)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., 2008 May 4 (806171335)

8)Rensing SA, Sugano S et al, The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants., Science, 319, 64-69 (2008) (801232011)

9)Yamamoto YY, Ichida H, Abe T, Suzuki Y, Sugano S, Obokata J., Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis., Nucleic Acids Res., 35, 6219-6226 (2007) (801231957)

10)Tsuritani K, Irie T, Yamashita R, Sakakibara Y, Wakaguri H, Kanai A, Mizushima-Sugano J, Sugano S, Nakai K, Suzuki Y., Distinct class of putative non-conserved promoters in humans: Comparative studies of alternative promoters of human and mouse genes., Genome Res., 17, 1005-1014 (2007) (801231948)

11)Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Sugano S, et al, The medaka draft genome and insights into vertebrate genome evolution., Nature, 447, 714-719 (2007) (801231941)

12)Sakakibara Y, Irie T, Suzuki Y, Yamashita R, Wakaguri H, Kanai A, Chiba J, Takagi T, Mizushima-Sugano J, Hashimoto SI, Nakai K, Sugano S., Intrinsic Promoter Activities of Primary DNA Sequences in the Human Genome., DNA Res., 14, 71-77 (2007) (801231936)

13)Takeda JI, Suzuki Y, Nakao M, Kuroda T, Sugano S, Gojobori T, Imanishi T., H-DBAS: Alternative splicing database of completely

sequenced and manually annotated full-length cDNAs based on H-Invitational., Nucleic Acids Res., 35(1):D104-109 (2007) (702102051)

14)Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S., Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs., Nucleic Acids Res., 35:D431-438, (2007) (702102044)

15)Yamamoto K, Ohishi M, Katsuya T, Ito N, Ikushima M, Kaibe M, Tatara Y, Shiota A, Sugano S, Takeda S, Rakugi H, Ogihara T., Deletion of angiotensin-converting enzyme 2 accelerates pressure overload-induced cardiac dysfunction by increasing local angiotensin II., Hypertension., 47:718-726 (2006) (702102110)

16)Osawa A, Kato M, Matsumoto E, Iwase K, Sugimoto T, Matsui T, Ishikura H, Sugano S, Kurosawa H, Takiguchi M, Seki N., Activation of genes for growth factor and cytokine pathways late in chondrogenic differentiation of ATDC5 cells., Genomics., 88(1):52-64 (2006) (702102105)

17)Khan A, Bohme U, Kelly KA, Adlem E, Brooks K, Simmonds M, Mungall K, Quail MA, Arrowsmith C, Chillingworth T, Churcher C, Harris D, Collins M, Fosker N, Fraser A, Hance Z, Jagels K, Moule S, Murphy L, O'Neil S, Rajandream MA, Saunders D, Seeger K, Whitehead S, Mayr T, Xuan X, Watanabe J, Suzuki Y, Wakaguri H, Sugano S, Sugimoto C, Paulsen I, Mackey AJ, Roos DS, Hall N, Berriman M, Barrell B, Sibley LD, Ajioka JW., Common inheritance of chromosome 1a associated with clonal expansion of *Toxoplasma gondii*., Genome Res., 16:1119-1125 (2006) (702102101)

18)Takeda J, Suzuki Y, Nakao M, Barrero RA, Koyanagi KO, Jin L, Motono C, Hata H, Isogai T, Nagai K, Otsuki T, Kuryshev V, Shionyu M, Yura K, Go M, Thierry-Mieg J, Thierry-Mieg D, Wiemann S, Nomura N, Sugano S, Gojobori T, Imanishi T., Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs., Nucleic Acids Res., 34:3917-3928 (2006) (702102058)

19)Cheong J, Yamada Y, Yamashita R, Irie T, Kanai A, Wakaguri H, Nakai K, Ito T, Saito I, Sugano S, Suzuki Y., Diverse DNA methylation statuses at alternative promoters of human genes

in various tissues., DNA Res., 13:155-167 (2006) (702102054)

21) Suzuki Y, Sugano S., Transcriptome analyses of human genes and applications for proteome analyses., Curr Protein Pept Sci., 7,147-163 (2006) (605251150)

22) Yamamoto N, Imai J, Watanabe M, Hiroi N, Sugano S., Yoshino G, Restoration of transforming growth factor-beta type II receptor reduces tumorigenicity in the human adrenocortical carcinoma SW-13 cell line., Horm Metab Res., 38,159-166 (2006) (605251147)

23) Kimura, K., Watanabe, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kushida, N., Yoneyama, T., Otsuka, R., Kanda, K., Yokoi, T., Kondo, H., Wagatsuma, M., Murakawa, K., Ishida, S., Ishibashi, T., Takahashi-Fujii, A., Tanase, T., Nagai, K., Kikuchi, H., Nakai, K., Isogai, T., and Sugano S., Diversification of transcriptional modulation: large-scale identification and characterization of purative alternative promoters of human genes, Genome Res., 16: 55-65, 2006 (603041657)

24) Otsuki T, Ota T, Nishikawa T, Hayashi K, Suzuki Y, Yamamoto J, Wakamatsu A, Kimura K, Sakamoto K, Hatano N, Kawai Y, Ishii S, Saito K, Kojima S, Sugiyama T, Ono T, Okano K, Yoshikawa Y, Aotsuka S, Sasaki N, Hattori A, Okumura K, Nagai K, Sugano S. Isogai T., Signal Sequence and Keyword Trap in silico for Selection of Full-Length Human cDNAs Encoding Secretion or Membrane Proteins from Oligo-Capped cDNA Libraries., DNA Res., 12: 117-126, 2005. (603041652)

25) Osada N, Hirata M, Tanuma R, Kusuda J, Hida M, Suzuki Y, Sugano S. Gojobori T, Shen CK, Wu CI, Hashimoto K., Substitution Rate and Structural Divergence of 5'UTR Evolution: Comparative Analysis Between Human and Cynomolgus Monkey cDNAs, Mol Biol Evol., 22: 1976-1982, 2005. (603021844)

26) Yamashita R, Suzuki Y, Sugano S. Nakai K., Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity., Gene, 350:129-136, 2005. (603021840)

[その他]

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、ヒトとマウスが多いが、今後他の生物を充実。なお、5' SAGEのHEK293とRamosの配列データに関しては、<http://5sage.gi.k.u-tokyo.ac.jp/>にて、生データを公開している。

DB名 : PrimerStation

URL : <http://ps.cb.k.u-tokyo.ac.jp/index.html>

PrimerStationは全ヒトゲノムに対して非常に高い特異性を持つ最適なプライマセットを計算するウェブサービス。本設計方法では高い精度を出すためにプライマの溶液中での会合率を用いる。会合率で厳格に計算するのは非常に計算時間がかかるため、これまでこのような物理化学モデルで計算し全ヒトゲノムに対してチェックをするようなウェブサービスはなかった。

6. 研究組織

(1) 研究代表者

菅野 純夫 (SUGANO SUMIO)

東京大学・大学院新領域創成科学研究科・教授

研究者番号 : 60162848

(2) 研究分担者

中井 謙太 (NAKAI KENTA)

東京大学・医科学研究所・教授

研究者番号 : 60217643

橋本 真一 (HASHIMOTO SHINICHI)

東京大学・大学院新領域創成科学研究科・特任准教授

研究者番号 : 00313099

山田 智之 (YAMADA TOMOYUKI)

東京大学・大学院新領域創成科学研究科・助教

研究者番号 : 50376604

(H17→H19)

(3) 連携研究者