

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5月29日現在

機関番号：14301

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21510215

研究課題名（和文） 非コード領域の保存配列モチーフの同定とそこに見られる多型の解析

研究課題名（英文） Identification of non-coding conserved motifs and their relationships with polymorphism

研究代表者

須山幹太（SUYAMA MIKITA）

京都大学・医学研究科・准教授

研究者番号：70452365

研究成果の概要（和文）：ゲノムアライメントを用い、生物種間で保存されている非コード領域の配列モチーフの同定を行った。モチーフとしてはスプライシング制御に関連したシス因子配列を主な対象とした。その結果、複数のモチーフが共存し、それらの組み合わせを介した高度な制御様式が存在することを明らかにした。転写因子結合部位モチーフへの拡張や多型との関連については今後の課題として、本研究遂行中に整備した研究環境およびプログラム群を用いて、引き続き研究していく。

研究成果の概要（英文）：By using genome sequence alignments, We have identified many non-coding motifs that are conserved among mammalian species. For these motifs, we mainly focused on the cis-regulatory elements that regulate splicing. As the results, we revealed the existence of a network of cis-regulatory elements to achieve complicated regulations of alternative splicing. Extending this methodology to the identification of transcription factor binding sites, and to analyze the relationships between these motifs and single nucleotide polymorphisms (SNPs) are still under way. The computational environments and programs developed in this research project should facilitate further progress of these analyses in future.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,500,000	450,000	1,950,000
2010年度	1,000,000	300,000	1,300,000
2011年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：複合新領域

科研費の分科・細目：システムゲノム科学

キーワード：ゲノム、バイオインフォマティクス、多型解析

1. 研究開始当初の背景

1.1. 比較ゲノム解析による DNA 配列モチーフの検出

研究開始当初までも数多くの DNA 配列モチーフの報告があったが、複雑な遺伝子発現制御を考えると、未だに発見されてい

ない制御領域が多数存在することが考えられた。それらの検出には比較ゲノム解析が有効であるといえ、それは現在においても変わっていない。これは、様々な動物ゲノム配列決定がなされ始めた当時ようやく可能になったアプローチである。具体的

には、複数種の脊椎動物ゲノム配列のアライメント（ゲノムアライメント）から遺伝子コード領域以外で保存している部位を収集し、さらにそれらの部位同士の比較から高頻度に現れるパターンがあれば、それが制御配列である可能性が高いとするものである。私はこの方法により、研究開始当時までも新規の転写因子結合部位やスプライス制御のためのシス因子の同定をおこなってきた（図1）。

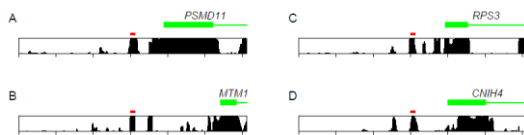


図1. 遺伝子上流に共通して見られる保存配列モチーフ CTACAACCTCCA。4つの遺伝子について表示してある(A-D)。モチーフ部分を赤、遺伝子を緑で示し、エクソンとイントロンをそれぞれ太線と細線で示してある。ボックス中は哺乳類のゲノムアライメント中での配列の保存度合を示し、保存が強い所が黒で表わされている。モチーフに相当する部位が局所的に保存しているのがわかる。ボックス下の目盛りは 100 塩基相当する。

1.2. ゲノム中に見られる擬陽性モチーフとの識別の重要性

上記の手順により DNA 配列モチーフが得られるが、これらのモチーフをゲノム上にマップしてみると非常に多くの擬陽性が出てしまう。これは配列モチーフが短く、場合によっては多少の変異を許容する弱いパターンであることに起因する。これまで、モチーフ検出については様々なプログラムが開発されてきているが、ゲノム中での擬陽性の識別についての研究はあまりなされていない。実際の転写因子などのトランス因子は真のモチーフと擬陽性のものを識別しているはずで、モチーフや配列プロファイルでは表現しきれない大局的な特徴を認識していると考えられる。私はこれまでの DNA 配列モチーフ検出の経験から、この擬陽性の識別が発現調節機構を解明する上での大きな課題であると考えられるようになった。

2.3. 疾患因子同定への応用

遺伝子の発現制御に関与する DNA 配列モチーフをゲノム中に同定することは、様々な応用的研究を促進するものである。例えば、発現制御領域での変異による疾患発症機構の解明といった医学的応用が考えられる。

遺伝統計学による病原因子の探索において、関連解析がよく用いられる。これは、ケース（患者）とコントロール（健常者）の間での多型の違いを統計的手法により比較

し、ケースに有意に見られる多型から疾患に関連した座位を検出しようとするものである。多型としては一般に SNP が用いられる。統計的に有意な SNP が遺伝子コード領域に存在した場合、その後の機能解析、すなわち予想される機能変化やその証明のための実験の組み立ては比較的容易だが、非コード領域に存在した場合、まずその機能の手がかりを得ることすら困難である場合が多い。このような場合、当該 SNP が DNA 配列モチーフ中に存在しているかどうかを確認することで、発現制御に関する異常による疾患である可能性を示唆することができる。

2. 研究の目的

ゲノム中には、遺伝子コード領域以外にも、最近解明が進んできたインシュレーター部位や転写因子結合部位など、発現制御に関わる重要な領域が多数存在する。それらの領域での変異は、遺伝子コード領域における機能性多型と同様に、遺伝的疾患に深く関与していることが考えられる。しかし、ゲノム中での発現制御領域に関する情報は限られており、病原因子となる座位の探索の上で、候補領域になりにくい。現在、様々な脊椎動物のゲノム配列決定が急速に進行していることから、それらを基にした比較ゲノム解析により、ゲノム中での発現制御領域についての知見を飛躍的に増加させることができる段階に来ている。特に、非コード領域中で生物種間に保存している領域は、発現制御などの重要な役割を担っていると考えられる。以上を踏まえ、本研究は、(1) 比較ゲノム解析により非コード領域中に見られる保存された DNA 配列モチーフの網羅的な抽出を行なうこと、(2) そこに見られる多型のうち、機能上の差異をもたらすような多型（機能性多型）を探索すること、の2点を大きな目的とした。

3. 研究の方法

本研究は、まずゲノムアライメントをもとに種間および種内での配列の保存から DNA 配列モチーフの検出を行なう。ゲノムアライメントは長大である上、生物種に特異的な反復配列やシンテニーの境界における断片化も非常に多く、扱うには技術を要する。そのため、まずゲノムアライメントから目的の配列部分を切り出すといった基本的な操作のためのプログラムを開発済した。

当初は主に転写因子結合部位を想定していたが、研究の進行に伴い、対象をスプライシングを制御するシス因子の同定とした。これは、モチーフが短くても揺らぎが少なく、検出がより容易であると考えられたこと、また、転写情報の蓄積にともない、選択的スプライシングによるアイソフォームの同定が

格段に進行したことによる。具体的には、UCSC のゲノムデータベースに登録されているすべてのヒト転写産物のデータをもとに、各遺伝子ごとに転写産物の構造を比較することで 1473 個の遺伝子について 1736 例のエクソンスキッピングを同定した。次にこれらのスキップされるエクソンの上流側および下流側のイントロン配列のゲノムアライメントを取得し、その中で進化的に保存している局所配列を探索した。スキップされるエクソン特異的に存在する配列モチーフを得るため、対照としてスキップされないエクソン周辺のイントロン配列を用いた統計検定をおこなった。

この解析で得られたモチーフの情報から、現在はエクソンスキッピングが観測されないが、モチーフの存在からエクソンスキッピングの可能性が示唆される例を複数得た。それらについては様々な組織由来の cDNA ライブラリを用いた RT-PCR 実験により、新たなエクソンスキッピングの有無を検証した。

4. 研究成果

スプライシングに関与する非コード領域に存在するシス配列モチーフについては、大きな成果が得られた。すなわち、それらは種間で保存しているものが多く、また、複数のモチーフが共存することで、それらの組み合わせを介した高度な制御様式が明らかとなった (雑誌論文②) (図 2)。

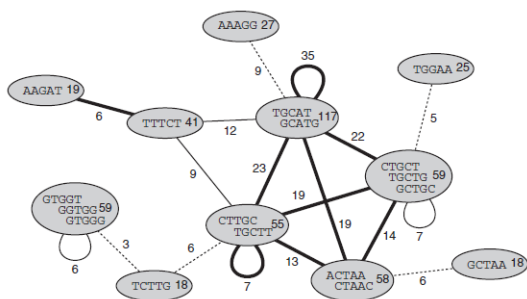


図 2. スプライシングに関連したモチーフのネットワーク。各ノードはモチーフを表しており、数字は観測された保存モチーフの数を示している。各エッジはモチーフの共存を表しており、数字は観測されたモチーフペアを示している。エッジの太さはモチーフペアの共存の統計的有意性を表している (太線: $p < 1.0 \times 10^{-5}$; 細線: $p < 1.0 \times 10^{-4}$; 破線: それ以外のモチーフをもっとも近いモチーフと結合)。

また、複数のモチーフが存在することを指標に、現在の転写産物データ中ではエクソンスキッピングが検出されていないが、その可能性を示唆する例を複数得た。それらについて様々な組織由来の cDNA ライブラリを用い

た RT-PCR により実験的な検証をおこなった。その結果、新たなスキッピング 3 例を実験的に証明することに成功した。これは今回得られたモチーフの共存情報が選択的スプライシングの予測へ応用できることを示している。

この研究成果は Nucleic Acids Res. 誌の Top 5% Article にも選ばれ、紹介記事とともに掲載された。この成果を得るにあたって、モチーフの統計的有意性を判別する方法を開発したが、この方法は、今後継続して研究を拡張したいと考えている、転写因子結合部位についての有意性判断についても適用可能である。

現在、スプライシングを制御する非コードモチーフの解析対象をエクソンスキッピングから排他的スプライシングに移し、さらなる研究を進めている。これまでのところ、cDNA 配列の比較解析から、ヒトにおいて約 50 例の排他的スプライシングの例を得ることに成功した。これまでもスプライシングのパターンのデータベースは存在したが、排他的スプライシングは分類が困難で、包括的な分類研究においても分類ミスが非常に目立っていた。本研究で得た排他的スプライシングの事例については、すべて目視により詳細な検討をおこなっており、これまでの分類に比べ非常に精度の高いデータセットであるといえる。このデータセットを元にして、排他的スプライシングを制御しているゲノム中の非コード因子の解析を進めている段階である。

本研究では、最終的にはゲノム DNA 配列モチーフ中に見られる機能性多型の検出を目指したが、機能と関連づいた多型の数の割合が、同定されている多型にくらべて圧倒的に少ないことから、網羅的な関連付けは今後の課題として残った。この研究課題により計算機環境およびプログラム開発が進み、また、モチーフの同定においても大きな前進が見られたことから、引き続き、多型との関連についての研究を進めていきたい。

本研究の遂行中に、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 (RNA-seq 法) や転写因子結合部位の同定 (ChIP-seq 法) において、格段の進歩が見られた。また、それに伴うデータの蓄積も著しい。研究遂行時にそれらのデータにも対処できるよう柔軟に軌道修正を行ってきた。よって本研究で得られた結果および開発してきた解析手法は、今後これらのデータ解析をしていく上でも非常に重要なものとなるといえる。実際、それらの次世代シーケンサーのデータ解析を精力的に進めつつあり、今後、本研究のさらなる発展が期待できる。

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 C. Ratanajaraya, H. Nishiyama, M. Takahashi, T. Kawaguchi, R. Saito, Y. Mikami, M. Suyama, M. Lathrop, R. Yamada, O. Ogawa, F. Matsuda. (2011) A polymorphism of the POLG2 gene is genetically associated with the invasiveness of urinary bladder cancer in Japanese males. *J. Hum. Genet.* **56**, 572-576. 査読有
DOI: 10.1038/jhg.2011.60.
- ② Suyama M, Harrington ED, Vinokourova S, von Knebel Doeberitz M, Ohara O, Bork P. (2010) A network of conserved co-occurring motifs for the regulation of alternative splicing. *Nucleic Acids Res.* **38**, 7916-7926. 査読有
DOI: 10.1093/nar/gkq705.

[学会発表] (計3件)

- ① 宮林香奈子, 嶋雄一, 馬場崇, 大竹博之, 佐藤哲也, 大川恭行, 須山幹太, 諸橋憲一郎 (2011, 12/13-16).
RNA-Seq 解析による胎仔型および成獣型ライディッシュ細胞特異的発現マーカー遺伝子の検索.
第34回日本分子生物学会年会, 横浜.
- ② 馬場崇, 大竹博之, 宮林香奈子, 嶋雄一, 佐藤哲也, 木村宏, 大川恭行, 須山幹太, 諸橋憲一郎 (2011, 12/13-16).
ChIP-Seq 法による Ad4BP/SF-1 の標的遺伝子の同定.
第34回日本分子生物学会年会, 横浜.
- ③ 佐藤哲也, 須山幹太 (2011, 12/13-16).
ChIP-Seq データを用いた核内受容体共役因子の探索.
第34回日本分子生物学会年会, 横浜.

6. 研究組織

(1) 研究代表者

須山 幹太 (SUYAMA MIKITA)
京都大学・医学研究科・准教授
研究者番号: 70452365