

転写ネットワーク発展の礎となったゲノム進化を同定

国立大学法人東海国立大学機構 名古屋大学大学院医学系研究科(研究科長・門松健治)・神経遺伝情報学(教授・大野欽司)の河地利彦(かわちとしひこ)医学部6年生、増田章男(ますだあきお)准教授らの研究グループは、脊椎動物に共通に保存されているC塩基の偏在が、液—液相分離を介した遺伝子発現制御の基盤となっていることを発見しました。

すべての生物で、ゲノムDNAはACGTの4種類の塩基から成り、タンパク質配列がACGT配列に置き換えて格納されています。生物は、大きなゲノム配列の変化を経て進化してきましたが、私たちヒトに至る過程で、どのような変化にどんな進化上の意義があったのかは、大きく謎に包まれています。

今回、タンパク質をコードするゲノム領域の中で、エクソン^{*1}と呼ばれる単位に着目して解析を行ったところ、C塩基が、生物進化とともに、とりわけ長い(多くの塩基数から成る)エクソンに偏在していくことを発見しました。このC塩基の偏在は、脊椎動物で特に顕著で、その配列を基に合成される蛋白は、天然変性領域と呼ばれる液—液相分離に必要な領域を形成し、RNA合成を司る転写因子群の機能発現を担っていました。生物は、このC塩基の偏在により、今の脊椎動物に見られる複雑で高度な転写ネットワークを獲得したと考えられます。さらに、一般的に長いエクソンほど、その転写・翻訳は困難になりますが、RNA代謝因子SRSF3の発現が、それを補い維持していることを同定しました。SRSF3は、癌遺伝子としても知られ、本研究成果は、生物進化研究のみならず様々な病態解明に寄与することが期待されます。

本研究成果は、科学雑誌「The EMBO Journal」に掲載されました(2021年10月4日付電子版)。

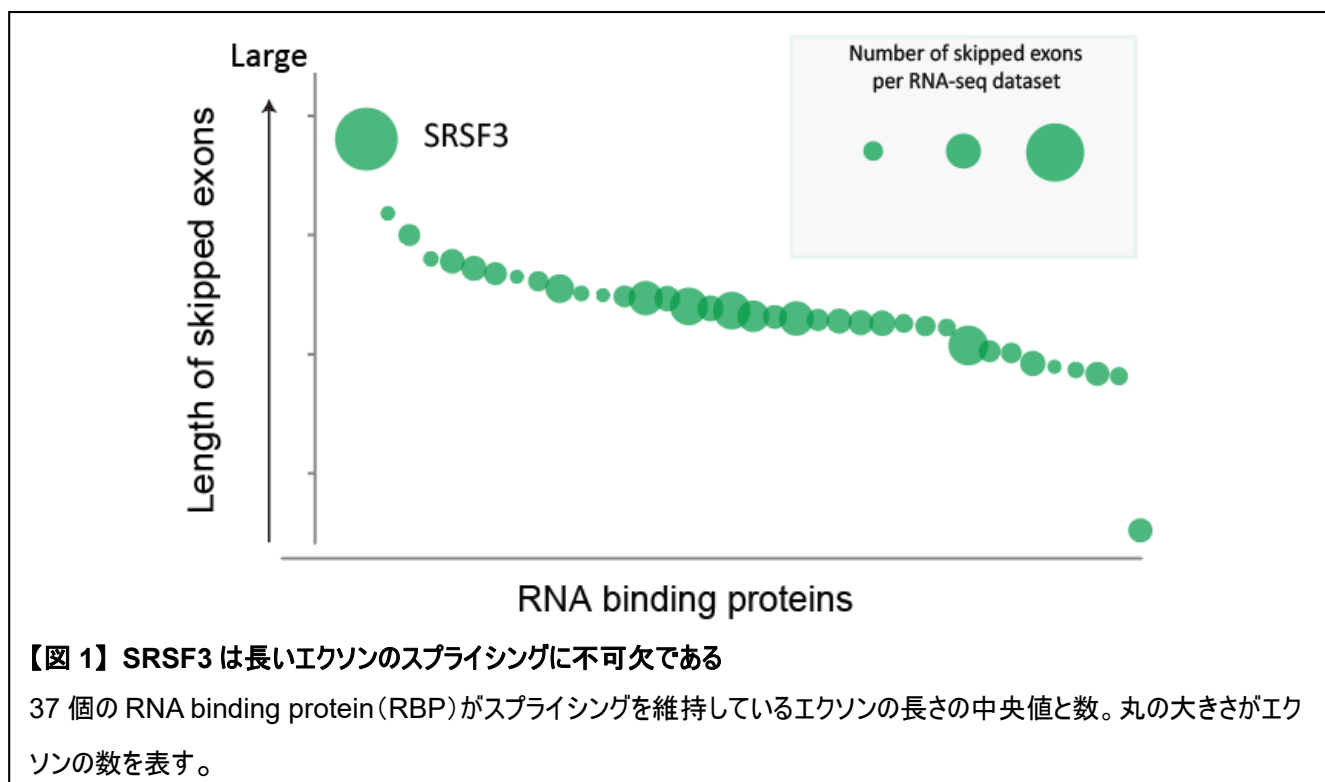
ポイント

- ゲノムに 3000 個以上ある長いエクソンのスプライシング^{※2}に必須な因子として SRSF3 を同定しました。
- 長いエクソンは、C 塩基に富み、これは RNA 代謝因子 hnRNP K の結合を誘導します。hnRNP K の結合はスプライシングを阻害しますが、SRSF3 が拮抗して働きます。結果、これら長いエクソンのスプライシングは保持され、この領域の蛋白への翻訳が保証されています。
- 長いエクソンにおける C 塩基の偏在は、プロリンやセリンなど特定の amino acid コードを増加させ、天然変性領域^{※3}とよばれる、液—液相分離を介した転写因子の相互活性を司る領域の形成を促しています。
- 長いエクソンへの C 塩基の偏在は、進化とともに進み、脊椎動物で特に顕著です。SRSF3 が長いエクソンのスプライシングを保証することで、転写因子の多様なネットワーク活動の発展が促され、脊椎動物の高度な生命活動が形作られてきました。

1. 背景

生物の遺伝情報は DNA から転写により mRNA へ、さらに翻訳により蛋白へと流れていきます。真核生物では、mRNA 合成時に、スプライシングと呼ばれるステップがあり、蛋白をコードしているエクソン同士が繋がり、間にあるイントロンが切り取られ捨てられることで蛋白翻訳のもととなる成熟 mRNA が形成されます。

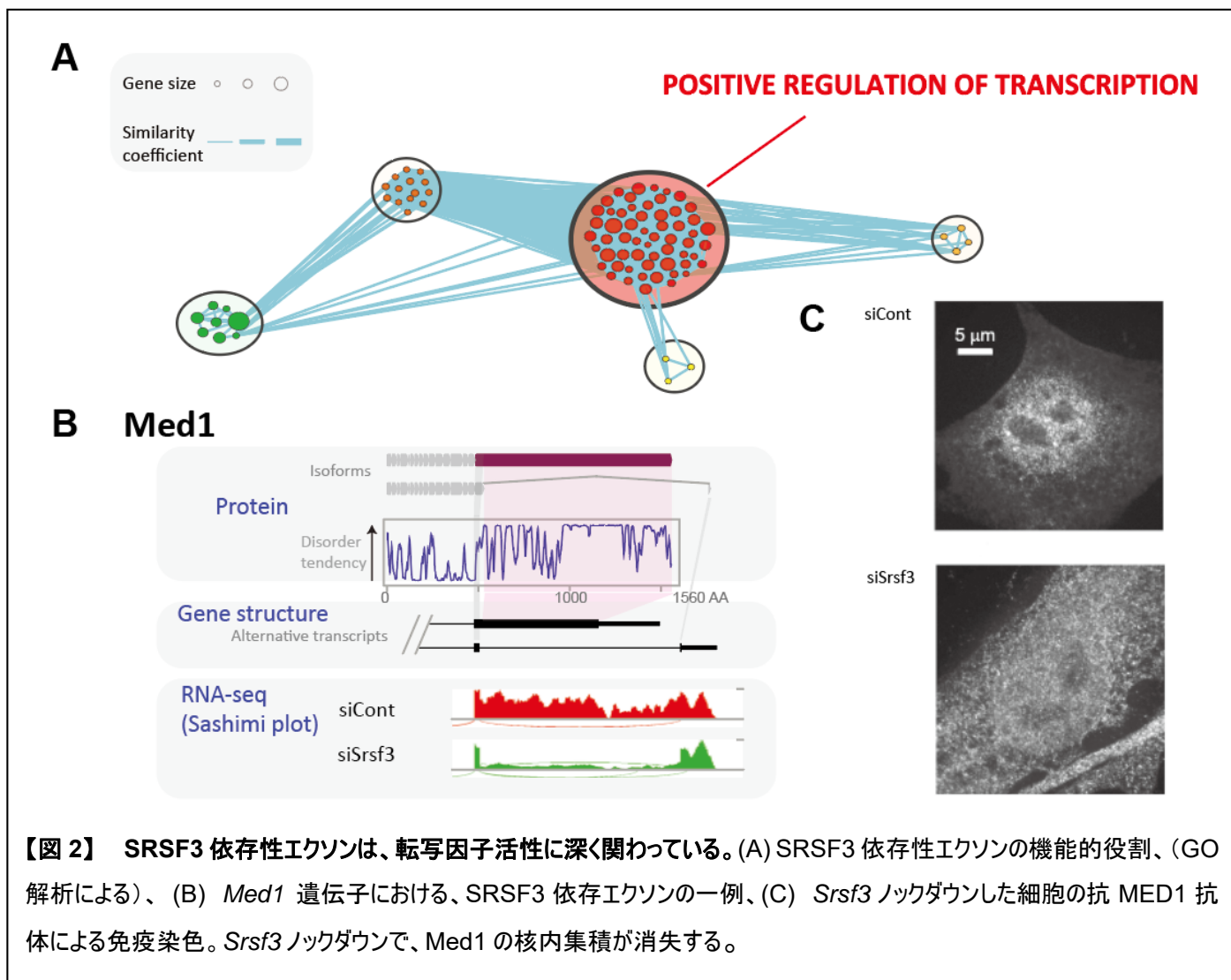
高等生物のエクソンは、通常 120 塩基程度（中央値）と小さく、ほとんどのエクソン（ヒトでは 95% 以上）が 300 塩基以下です。実験的には、エクソンはサイズが大きくなるほどスプライシングされ難くなることが知られています。しかし、実際には、多くの大きなエクソンが進化的に保存されており、きちんとスプライシングされています。これら大きなエクソンは、どうやってスプライシングが行われ、何のために進化的に保存されているのでしょうか？本研究で、これらの疑問の解明を目指しました。



2. 研究成果

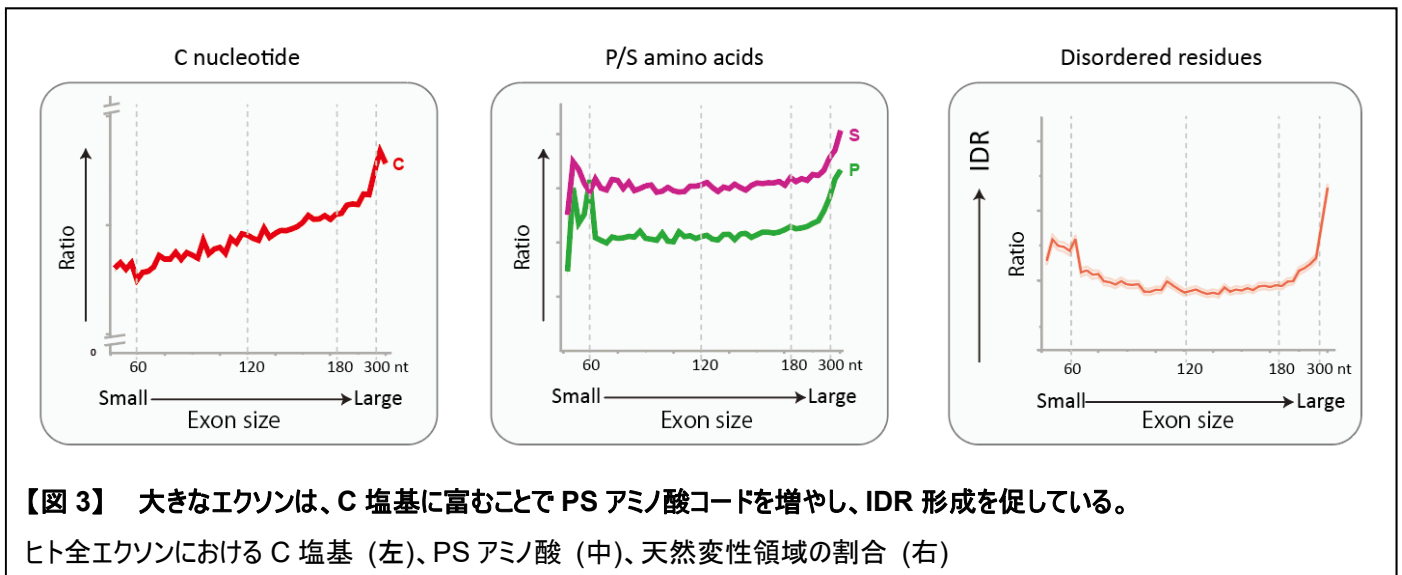
公共データベースに登録されている、RNA 結合蛋白 (RBP) をノックダウン/アウトした 800 個以上の RNA-seq データを解析し、SRSF3 が突出して大きなエクソンのスプライシングに必要であることを見出しました (SRSF3 の欠失によりスキップされるエクソン長: 中央値 190 塩基、他の RBP: 中央値 159~90 塩基) (図 1)。

ENCODE に登録されている eCLIP データセットを解析したところ、SRSF3 依存エクソンには hnRNP K の集積が顕著で、*Srsf3* および *Hnrnpk* のノックダウン RNA-seq 解析の結果、hnRNP K はこれらエクソンのスプライシングを阻害する活性を持つものの、SRSF3 がこの作用を完全に打ち消してスプライシングを維持していることが示唆されました (図 2)。



【図 2】 SRSF3 依存性エクソンは、転写因子活性に深く関わっている。(A) SRSF3 依存性エクソンの機能的役割、(GO 解析による)、(B) *Med1* 遺伝子における、SRSF3 依存エクソンの一例、(C) *Srsf3* ノックダウンした細胞の抗 MED1 抗体による免疫染色。*Srsf3* ノックダウンで、Med1 の核内集積が消失する。

遺伝子オントロジー (GO) 解析の結果、SRSF3 依存エクソンを持つ遺伝子は、ほとんどが転写制御因子でした (図 2A)。これらエクソンは、蛋白間相互作用の中核をなす液—液相分離に必須な天然変性領域 (IDR) をコードしていました。



ヒトゲノムに存在するすべてのエクソンを解析したところ、大きなエクソンには C 塩基が顕著に存在することが判明しました(図 3 左)。C 塩基の増加は、プロリンやセリンのアミノ酸コードの濃縮を引き起こし(図 3 中)、これらは天然変性領域の形成につながっていました(図 3 右)。進化解析では、この C 塩基の偏在は、特に脊椎動物の長いエクソンに広範囲に渡って認められます。進化の過程で、プロリンとセリンをコードしながら C 塩基が長いエクソン特異的に増加したことで、転写因子の IDR が形成されました。C 塩基の増加は、hnRNP K などのスプライシング阻害因子の活性化をもたらし、長いエクソンのスプライシングを困難にします。そこで、これらのスプライシングを維持するために SRSF3 が出現したと考えられます。

3. 今後の展開

我々ヒトを含む脊椎動物は、大きなゲノム変化の結果として、それまでの生物とは一線を画す進化を示しています。今回の研究は、その変化としてエクソンレベルでの天然変性領域の獲得があったことを示しており、今後の進化研究に大きなインパクトを与えるものと予想されます。また、この機構維持に必須な SRSF3 は、ヒトの多くのがん、特に卵巣がんや乳がん、肺がんにおいて過剰発現し、がん細胞増殖に必須のタンパク質であることが知られています。今回の発見は、SRSF3 が広範な転写活動の維持に必須であることを示唆しており、これら疾患の病態解明に寄与することが期待されます。

4. 用語説明

※ 1 エクソン

真核生物のほとんどの遺伝子において、タンパク質の情報に相当する部分はゲノム DNA 中に分断されて存在します。遺伝情報がコードされている領域がエクソン、遺伝情報がコードされていない領域がイントロンと名付けられています。

※ 2 スプライシング

DNA から転写された RNA の一部を切り取り、エクソン同士をつなげる機構。スプライシングによってつながられた RNA が成熟 mRNA として細胞質に運ばれ翻訳されることで、蛋白が合成されます。

※3 天然変性領域

細胞中で三次元立体構造をとらずに伸びた構造をとる領域を指しています。荷電や極性などが偏った、限られたアミノ酸組成で構成されることが特徴です。近年、この領域は液—液相分離と呼ばれる膜に依存しない機能性分画形成を、促進する領域であることが明らかとなり、注目されています。

5. 発表雑誌

掲雑誌名: The EMBO journal

論文タイトル: Regulated splicing of large exons is linked to phase separation of vertebrate transcription factors

著者: Toshihiko Kawachi^{1,2}, Akio Masuda^{1,2,*}, Yoshihiro Yamashita¹, Jun-ichi Takeda¹, Bisei Ohkawara¹, Mikako Ito¹, Kinji Ohno¹

所属: ¹Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan. ²These authors equally contributed to this work.

* Corresponding author

DOI: <https://www.embopress.org/doi/10.15252/emboj.2020107485>

English ver.

https://www.med.nagoya-u.ac.jp/medical_E/research/pdf/EMBO_211004en.pdf