

In silico 解析と *in vitro* 解析による RNA スプライシング機構の研究

●大野 欽司

名古屋大学大学院医学系研究科神経遺伝情報

<研究の目的と進め方>

ヒトゲノムプロジェクトが完了し、ヒトは22,000個という限られた数の遺伝子から組織特異的・発達段階特異的に splicing trans-factors を発現させ、alternative splicing を行い、10万種類以上と予想される多様なタンパク質を実現していることが明らかになり、splicing の生理・病態機構の解明は従前にも増して重要になってきている。

本研究の目的は、ヒトゲノム上に annotation をされた全 transcriptome の *in silico* 解析と培養細胞を用いた *in vitro* 解析により各 splicing cis-element 内、及び、各 cis-elements 相互の塩基配列の法則を解明することである。さらに、各 cis-element 内、及び、cis-elements 間の法則を明らかにすることにより degeneracy の統合的な代償機構を解明し、splicing 異常を起こす遺伝子変異を予測する精度の高い汎用アルゴリズムを構築することである。

1. Splice donor site の予測アルゴリズム

U1 snRNA が結合をする splice donor sites における遺伝子変異が splicing 異常を起こすことが従来より数多く報告されてきているが、現在でも見落とされている変異が多く存在する (personal observations)。本研究の目的は、splice donor sites における塩基間相互の関係を明らかにし、splicing 変異を効率よく予測するアルゴリズムの構築を行うことである。さらに、構築したアルゴリズムのウェブサービスを開始することである。

2. ヒト branch point コンセンサス塩基配列の決定

ヒトを含む高等動物の branch point コンセンサス配列は degenerative であるが、degeneracy が全く存在しない yeast の branch point 配列に基づいて予測アルゴリズムが作られている。自験例の検討によるとこのアルゴリズムは不正確である。またひとつのイントロンに複数の branch points が使われることが多く、branch point sequence の degeneracy の tolerance の定量化と正確な branch point 予測アルゴリズムは branch point sequence を破断する遺伝子変異の研究に重要である。本研究では、全 transcriptome 解析・lariat RT-PCR 法・ribonuclease protection assay 法によりヒトにおける branch point コンセンサス配列の決定を行う。

3. Aberrant splicing を惹起するエクソン 5' 末端遺伝子変異の予測アルゴリズムの構築

Exon +1 位変異による aberrant splicing は従来3例しか報告されておらず見逃されている例が多いと想定される。従来ほとんど研究が行われていない U2AF65、及び、U2AF35 が結合をする intron 3' 末端と exon 5' 末端の塩基間相互の関係を明らかにし、AG-dependent splice sites と AG-independent splice sites の判別アルゴリズムを構築する。これらの解析結果に基づき、exon +1 位を含む exon 5' 末端領域の遺伝子変異の中で、splicing 異常を起こすものを検出するためのアルゴリズムを構築する。

これらの解析を組み合わせることで疾患関連遺伝子変異が splicing 異常を起こす機構の解明と、splicing 異常を惹き起こす遺伝子変異を予測する汎用アルゴリズムの構築を試みる。

<研究開始時の研究計画>

1. Splice donor site の予測アルゴリズム

Splice donor site の exonic positions -2 位と -1 位において splicing 異常を示すことが知られている遺伝子変異を minigenes に導入し、さらに、artificial mutations を導入し、HEK293 細胞を用いて splicing 異常を再現する。同時に、ヒトゲノム上に annotation をされた全 transcriptome 情報を用い splice donor sites の exonic position -3 から intronic position +6 の塩基配列を抽出し、各塩基間相互の法則を見出す。これらの解析結果に基づく各種パラメータを用いたヒト splice donor sites における splicing 変異予測アルゴリズムを各種構築する。さらに多数例の遺伝子変異のミニジーンによる解析、及び、既報告の exonic and intronic splicing mutations を用いてアルゴリズムの検定を行い、最適パラメータと最適アルゴリズムの構築を行う。従来よりも正確なアルゴリズムの構築に成功した場合にはウェブサービスプログラムを研究コミュニティに提供する。

2. ヒト branch point コンセンサス塩基配列の決定

ヒトにおける branch point コンセンサス配列を実験結果に基づいて決定を行うために house keeping genes を対象に 100ヶ所の branch points を lariat RT-PCR により同定を目指す。House keeping genes は各種細胞において高発現であり、intron が短く、alternative splicing を受ける exon が少ないために本研究の対象遺伝子として最適であると考えられる。これらの結果を training dataset として、branch point 予測アルゴリズムを構築する。ヒト全ゲノムを対象にこの予測アルゴリズムによる branch points の予測を行い、予測結果の検定を lariat RT-PCR により複数の遺伝子の複数の introns を用いて行う。

House keeping genes を用いたヒト branch point コンセンサス配列の解析により、branch point の位置、ならびに、branch point の遺伝子変異のスプライシングに与える影響を予測するアルゴリズムを構築する。信頼できるアルゴリズムの構築に成功した場合にはウェブサービスプログラムを研究コミュニティに提供する。

3. Aberrant splicing を惹起するエクソン 5' 末端遺伝子変異の予測アルゴリズムの構築

Spliceosome の構築に intron 3' 末端の AG が必要な AG-dependent splice sites と不必要な AG-independent splice sites が存在する。AG-dependent splice sites においては、intron 3' 末端の AG に加えて exon 5' 末端の G が U2AF35 の結合に重要であり、exon 5' 末端の G の遺伝子変異が aberrant splicing の原因になるという仮説を立て実証を行う。FECH と GHI では exon 5' 末端の G の変異が exon skipping を起こすが、LPL と HEXA では起こさない。それぞれの遺伝子の minigene を作成し、各種 mutagenesis を行い exon skipping を調べることにより exon 5' 末端遺伝子変異が aberrant splicing を惹起するのに必要な cis-elements の配列要素を決定する。

エクソン 5' 末端遺伝子変異のスプライシングに与える影響を

予測するアルゴリズムを構築する。信頼できるアルゴリズムの構築に成功した場合にはウェブサービスプログラムを研究コミュニティに提供する。

<研究期間の成果>

1. Splice donor siteの予測アルゴリズム

ヒトゲノムの annotation 解析にて、1779,917ヶ所の GT dinucleotide を持つ splice donor sites を抽出した。これらの解析にて、exon 3' 末端3塩基に U1 snRNA と相補でない塩基がひとつでも存在する場合には、U1 snRNA は exon 領域に結合をすることを諦め intron 領域のみに結合をする傾向があることが判明した。つぎに、各 splice donor sites のヒトゲノムにおける出現頻度が splicing signal 強度を推定するよい指標であることを見出した。ミニジーンを使った実験結果を training dataset として用い、出現頻度の対数指標である SD-Score を使った splicing 異常予測アルゴリズムを構築した。さらに多数のミニジーンを使った実験結果、さらに、論文上に報告された splicing 変異を validation dataset に用い検証を行ったところ、sensitivity = 97.5%、specificity = 94.6% という精度を得た。この手法により、179,917ヶ所の splice donor site に一塩基置換を導入する simulation を行ったところ、exonic position -3, -2, -1 における変異のそれぞれ 38%, 89%, 97% が splicing 変異であると予測された。この予測に基づき、exonic position -3 の2つの遺伝子変異が splicing mutation であることをミニジーンを使って証明できた。ひとつの遺伝子変異については、患者サンプルを用いた解析にて、患者固体においても splicing 異常が起きていることを証明した。

これらの成果を支援班のサポートにより SD Score アルゴリズムのウェブサービスを開始した。

2. ヒト branch point コンセンサス塩基配列の決定

ヒト 20 種類の house keeping genes の 52 の introns を対象に 367 個の lariat clones を解析した。367 個の clones のうち branch point に misincorporated nucleotides を認めたものは 181 clones で、これらを用いて解析を行った。残りの 186 個のクローンは lariat RT-PCR において branch point がスキップをした可能性を否定できなかったため解析対象としなかった。Branch points の塩基は、92.3% A, 3.3% C, 1.7% G, 2.8% U であった。解析の結果、ヒトの branch point コンセンサス配列は yUnAy であることが判明した。Branch points の 83% は -34 位から -21 位に存在した。Polypyrimidine tract は branch point の下流 4 から 24 塩基に存在した。ヒトの branch point コンセンサス配列は、予想されるよりも degenerative であり、branch point は他の splicing cis-element(s) と同時に認識をされている可能性が高いと思われた。

3. Aberrant splicing を惹起するエクソン5'末端遺伝子変異の予測アルゴリズムの構築

ヒトゲノム上に存在するすべての splice acceptor sites の解析にて exon 5' 末端の塩基が G の時 polypyrimidine tract が短いことが明らかになった。つまり、polypyrimidine tract が長い intron は AG-independent であり、spliceosome assembly の第1段階において U2AF35 を必要としないために、exon 5' 末端の G を必要としないと推察された。

次に Exon 5' 末端の G の他の塩基への変異が exon skipping を起こす FECH, GHI, EYA1 遺伝子を用いて exon skipping を minigenes にて培養細胞で再現を行った。系統的に polypyrimidine tract を延長させることにより、FECH, GHI, EYA1 遺伝子でそれぞれ 13, 15, 10 塩基以上の polypyrimidine stretch を導入することにより splicing が正常化することを見出した。

polypyrimidine stretch を徐々に短くすることにより exon skipping の割合が増強した。一方、遺伝子変異が exon skipping を起こさない LPL, HEXA 遺伝子においては polypyrimidine tract を徐々に短縮させることにより、両者とも正常な splicing のためには 10 塩基以上の polypyrimidine stretch が必要であることが判明した。さらに polypyrimidine stretch を短くすることにより exon skipping の割合が増強した。

Exon skipping を起こす minigene constructs においては U2AF65 との in vitro 結合能が低下していることを示した。U2AF65 の siRNA によるノックダウンを試みたが U2AF65 は細胞の生存に必須の分子のために効率よくノックダウンされた細胞は死滅し、U2AF65 の効率的なノックダウンは不可能であった。一方、U2AF35 のノックダウンには成功し、AG-dependent 3' splice sites においては U2AF35 のノックダウンにて exon skipping が誘導されることを明らかにできた。

<国内外での成果の位置づけ>

1. Splice donor siteの予測アルゴリズム

Splice donor sites の遺伝子変異のスパライシングに与える影響を予測するアルゴリズムには Senapathy らの consensus value と Rogan らの information contents が従来報告をされている。Consensus value を用いて、我々の dataset の解析を行ったところ、sensitivity = 97.5 と我々と同様の感度を示したが、specificity = 89.2% と低い値であった。また、Information contents を用いて同様の解析を行ったところ、sensitivity = 80.1%, specificity = 94.6% という値であり、我々のアルゴリズムには及ばなかった。今後、我々の SD-Score アルゴリズムが、5' splice site における splicing 遺伝子変異を予測する標準的な手法になると期待される。

2. ヒト branch point コンセンサス塩基配列の決定

ヒト branch point コンセンサス塩基配列の決定プロジェクトでは、in vitro 実験結果に基づくコンセンサス配列を示すことができた。内外の splicing 研究者に本プロジェクトの成果を説明したところ、「分子生物学の教科書の記述を変える研究」との評価を受けている。また、mammalian branch point sequence の degeneracy が高いことは、個別の遺伝子の研究から各研究者が想像しており、本研究はその論理的な根拠を与えているとの評価を得ている。

3. Aberrant splicing を惹起するエクソン5'末端遺伝子変異の予測アルゴリズムの構築

Exon 5' 末端の点変異が splicing 異常を起こし得ることは、疾患関連遺伝子変異の研究者の多くが予想さえていない病態であり、その予測アルゴリズムの構築はヒト疾患関連遺伝子変異研究において重要なテーマになると期待をされるが、まだ効率のよいアルゴリズムの構築には至っていない。

<達成できなかったこと、予想外の困難、その理由>

1. Splice donor siteの予測アルゴリズム

同様のアルゴリズムが2種類報告をされている。Consensus value を用いて、我々の dataset の解析を行ったところ、sensitivity = 97.5 と我々と同様の感度を示したが、specificity = 89.2% と低い値であった。また、Information contents を用いて同様の解析を行ったところ、sensitivity = 80.1%, specificity = 94.6% という値であり、我々のアルゴリズムには及ばなかった。今後、我々の SD-Score アルゴリズムが、5' splice site における splicing 遺伝子変異を予測する標準的な手法になると期待される。

2. ヒト branch point コンセンサス塩基配列の決定

ヒト branch point コンセンサス塩基配列は yUnAy という 6 bit の情報しかない予想されるよりもはるかに degenerative な配列で

あった。下流の polypyrimidine tract の情報と統合して branch point の位置を決定する試みを行ったが、polypyrimidine tract は情報量が多いものの位置情報に乏しく、信頼性のあるアルゴリズムの構築が困難であった。

3. Aberrant splicing を惹起するエクソン5'末端遺伝子変異の予測アルゴリズムの構築

詳細な解析を行ったエクソンごとに正常なスプライシングに必要な polypyrimidine tract 長が異なり、唯一の基準を示すことができなかった。スプライシングシス因子のシグナル強度はイントロン 3' 末端のみで決定されるわけではなくすべてのシス因子シグナル強度の統合的な解析が必要と思われる。

<今後の課題、展望>

すべてのスプライシングシス因子を統合して定量的に評価を行うアルゴリズムの構築を目指したい。また、そのようなアルゴリズムのウェブサービスを行いたいと願っている。

<研究期間の全成果公表リスト>

1) 論文/プロシーディング

1. 0801291427

Sahashi K, Masuda A, Matsuura T, Shinmi J, Zhang Z, akeshima Y, Matsuo M, Sobue G, Ohno K. In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites. *Nucleic Acids Research*, 35, 5995-6003 (2007).

2. 0801291432

Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Ishida M, Shinmi J, Yatsuya H, Qiao S, Takahashi M, Ohno K. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids Research*, 35, e123 (2007).

3. 0801291434

Saito T, Amakusa Y, Kimura T, Yahara O, Aizawa H, Ikeda Y, Day JW, Ranum LPW, Ohno K, Matsuura T. Myotonic dystrophy type 2 in Japan: ancestral origin distinct from Caucasian families. *Neurogenetics*, 9: 61-63 (2008).

4. 0805161026

Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 2008, 36: 2257-2267.

5. 0901161125

Shen X-M, Fukuda T, Ohno K, Sine SM, Engel AG. Congenital myasthenia-related AChR delta subunit mutation interferes with intersubunit communication essential for channel gating. *J Clin Invest* 2008, 118: 1867-1876.

6. 0901161134

Masuda A, Shen XM, Ito M, Matsuura T, Engel AG, Ohno K. hnRNP H enhances skipping of a nonfunctional exon P3A in *CHRNA1* and a mutation disrupting its binding causes congenital myasthenic syndrome. *Hum Mol Genet* 2008, 17: 4022-4035.

7. 0901161138

Bian Y, Masuda A, Matsuura T, Ito M, Okushin K, Engel AG, Ohno K. Tannic acid facilitates expression of the polypyrimidine tract binding protein and alleviates deleterious inclusion of *CHRNA1* exon P3A due to an hnRNP H-disrupting mutation in congenital myasthenic syndrome. *Hum Mol Genet*

2009, 18: 1229-1237.

2) データベース/ソフトウェア

- 5' splice site の遺伝子変異のスプライシングに与える影響を 97.1% の感度と 94.7% の特異度で判別をする SD-Score algorithm (登録番号 0801291427. Sahashi K, Masuda A, Matsuura T, Shinmi J, Zhang Z, akeshima Y, Matsuo M, Sobue G, Ohno K. In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites. *Nucleic Acids Res*, 35: 5995-6003, 2007) を 2008 年度情報支援班の支援によりウェブサービスを開始した。

http://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html

- 8 種類の既報告 siRNA 設計パラメータと我々が開発をした iScore (登録番号 0801291432. Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Ishida M, Shinmi J, Yatsuya H, Qiao S, Takahashi M, Ohno K. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids Res*, 35: e123, 2007) の合計 9 種類のパラメータに加えて 5 種類の熱力学的なパラメータを計算するウェブサービスを 2008 年度情報支援班の支援により開始をした。従来の siRNA 選択ツールが 1 種類のパラメータのみを計算するのに比して、9 種類を同時に比較検討することにより効率よく siRNA の設計が可能になっている。

http://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html