

平成 30 年 6 月 27 日現在

機関番号：34204

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00411

研究課題名(和文)機能未知スプライシングアイソフォームの機能部位予測法の開発

研究課題名(英文)Development of a method for predicting functional sites in function-unknown splicing isoforms

研究代表者

塩生 真史 (Shionyu, Masafumi)

長浜バイオ大学・バイオサイエンス学部・准教授

研究者番号：30345847

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：現在、選択的スプライシングにより作られるスプライシングアイソフォームの配列情報が多数知られている。しかし、その多くは機能が不明であり、そもそも機能性があるかどうか議論になっている。そこで、スプライシングアイソフォームが機能を持つかどうかを推定する方法を開発したところ、これまでに議論されているよりも多くのもの機能を持つことが示された。また、機能があると推定されるものについてその機能部位を推定する方法の開発を行った。その結果、開発した低分子結合部位予測法は以前のものよりも大幅に精度が向上していた。さらに、既存法に比べて良い精度で結合する低分子を判定できる手法を開発することができた。

研究成果の概要(英文)：At present, many sequence data of splicing isoforms produced by alternative splicing are known. However, most of these splicing isoforms are function-unknown. In addition, it is argued whether they are functional or not. Thus, we developed a prediction method for functionality of splicing isoforms. Using our method, 70% of splicing isoforms, of which protein expression were not confirmed, were predicted to be functional, and this value is larger than one shown before. Moreover, we developed a method for predicting functional sites of splicing isoforms predicted to be functional. The method we developed showed higher accuracy for predicting small molecule-binding sites than that of our previous method. And we developed a method for predicting a small molecule ligand of a target protein that showed better performance than that of conventional one.

研究分野：情報生物学

キーワード：選択的スプライシング 機能未知タンパク質 機能部位予測 低分子化合物 SNV データベース 機械学習

## 1. 研究開始当初の背景

ヒトでは、選択的スプライシングと呼ばれる現象により1遺伝子当たり平均3種類以上の mRNA が生じており、知られているヒトタンパク質のアミノ酸配列数は 80,000 を超えている。このような選択的スプライシングによる遺伝子産物の多様化が、限られた数の遺伝子から生物学的複雑さを生み出す要因の1つと考えられている。しかし、各遺伝子から選択的スプライシングによって作られる産物(スプライシングアイソフォーム)のうち、主要なものについては機能が明らかになっていても、それ以外のマイナーなものについては、実験的な検証が行われておらず、どのような機能を持つか不明なことが多い。そのため、多数知られているマイナーかつ機能未知なスプライシングアイソフォームが、実際に生物学的複雑さに寄与するかについてはまだ明らかになっていない。

研究代表者は、ヒトスプライシングアイソフォームの情報生物学的な解析により、8割以上のマイナーなスプライシングアイソフォームにおいて、タンパク質立体構造形成に重要である疎水性コアや、他分子との相互作用部位の一部が欠損していることを示し、このようなスプライシングアイソフォームが生化学的な機能を失っている可能性を初めて明らかにした(Yura & Shionyu et al., 2006)。その後、他の研究グループにより類似の手法による解析が行われ、マイナーなスプライシングアイソフォームの多くがスプライシング過程のゆらぎによって偶然に作られた、機能的には意味がないものであるという議論がなされた(Melamed & Moul t, 2009)。

このような解析結果は、選択的スプライシングが生物学的複雑さを生み出す要因の1つであるという考え方と矛盾する。一方で、これらの解析はスプライシングアイソフォームの生化学的な機能に焦点が当てられ、生物学的な機能に対しての寄与はほとんど考慮されていない。そこで研究代表者は、生化学的な機能との関連が深い機能部位や立体構造のみには依存せずヒトスプライシングアイソフォームの機能性を推定できる方法が必要であると考え、機能性のあるスプライシングアイソフォームが満たすと考えられる必要条件に基づいて機能的なスプライシングアイソフォームかどうかを判断する機能性推定法を開発した。開発した手法を疎水性コアが欠失していて安定な立体構造が形成できないと考えられるスプライシングアイソフォームに適用したところ、約 27%のスプライシングアイソフォームについて機能性があると推定された。このことから、疎水性コアや既知の機能部位の欠損だけでは、そのスプライシングアイソフォームに機能的な意味がないと判断することはできないことが明らかになった。

この時開発した機能性推定法は、「タンパク質レベルでの発現が確認できる」、「mRNA レベルで発現調節がされている」、および、「異なる生物種間でのスプライシングイベントの保存されている」という条件のいずれかを満たすかどうかを判定しているだけであり、機能的なスプライシングアイソフォームが取りこぼされている可能性が高い。また、スプライシングアイソフォームが何らかの機能を持つかどうかを判定するだけであり、そのスプライシングアイソフォームが持つ具体的な機能は推定できていない。

## 2. 研究の目的

### (1) スプライシングアイソフォームの機能性推定法の開発

研究代表者が開発したスプライシングアイソフォームの機能性推定法に、スプライシングアイソフォームの様々な特徴を取り入れることで、マイナーな機能未知のスプライシングアイソフォームが生体内で機能を持つかどうかをより精度良く推定できるようにする。

### (2) 立体構造情報に基づく低分子結合部位及び結合低分子予測法の開発

機能性推定法により機能を持つと推定されたスプライシングアイソフォームについて、その機能を推定するために、これまでに研究代表者が開発してきたタンパク質の立体構造情報にもとづく低分子結合部位の予測法を発展させて、低分子結合部位だけでなく、結合する低分子も推定できるようにする。

### (3) シングルヌクレオチドバリエーション(SNV)に基づく機能部位の解析

疾患の原因ともなり得る SNV が、タンパク質上の機能部位とどのような関係にあるかを明らかにすることで、立体構造情報が得られないスプライシングアイソフォームについても機能部位を推定できるようにする。

## 3. 研究の方法

### (1) スプライシングアイソフォームの機能性推定法の開発

まず、個々のスプライシングアイソフォームについて、具体的にどのような機能が実験的に調べられているかについて、公共データベースである UniProt/Swiss-Prot や文献情報を元に調査した。

また、機能が見られるスプライシングアイソフォームがどのような特徴を持つかを明らかにするため、これまでの機能性推定法で用いていた指標に加えて、新たに球状ドメインを形成する領域および天然変性領域の増減、機能ドメインおよび膜貫通部位の増減、疎水性コアへの影響の有無、選択的スプライシングの種類などの指標を、各スプライシングアイソフォームについて解析した。

このようにして得られたスプライシング

アイソフォームの特徴を、機械学習により学習することで、機能が調べられているスプライシングアイソフォームと類似の特徴を持つものを機能性のあるものと推定することにした。

ここで、一般的な教師あり学習の手法では、陽性（本研究の場合は「機能がある」）と陰性（本研究の場合は「機能がない」）の特徴の違いを学習するが、多くのスプライシングアイソフォームは機能があるかどうか不明であり、一般的な機械学習の枠組みでは学習できない。そこで、半教師あり学習の一種である Positive-Unlabeled (PU) アルゴリズム (Elkan & Noto, 2008) を Random Forest 法と組み合わせることで、機能的なスプライシングアイソフォームの特徴の学習を行った。

## (2) 立体構造情報に基づく低分子結合部位及び結合低分子予測法の開発

研究代表者はこれまで、低分子結合部位に存在するアミノ酸残基の統計情報にもとづいて、低分子との結合部位を予測する方法 (Propensity of Ligand Molecule-binding Sites, 以下 ProLMS) を開発してきた。この ProLMS で得られたスコアに加えて、タンパク質表面におけるポケット状の構造上での位置、および、部位の進化的な保存度を低分子結合部位の特徴量として用いて、機械学習 (Random Forest 法) により学習することで ProLMS よりも高精度で目的の低分子結合残基を予測できるようにした。さらに、予測された低分子結合残基が空間的に近接して存在しているかどうかを判定することで、標的とするタンパク質にその低分子に結合するかを予測できるようにした。

## (3) シングルヌクレオチドバリエーション (SNV) に基づく機能部位の解析

数万人規模の健常人全エクソンバリエーションデータベースである ExAC から SNV データをダウンロードし、これら SNV のデータを、ヒトの mRNA 塩基配列、および、タンパク質のアミノ酸配列にマッピングすることで、SNV と機能部位の位置関係を解析できるシステムの構築を行った。

また、スプライシングアイソフォームで共通に使われているエクソンと、選択的に使われるエクソンの間で一塩基変異 (SNV) の分布の比較を行った。

## 4. 研究成果

### (1) スプライシングアイソフォームの機能性推定法の開発

UniProt/Swiss-Prot データベースにおいて、ヒトのマイナーなスプライシングアイソフォームの中で、何らかの機能に関する情報が調べられているものは、約 1300 個であった (図 1)。

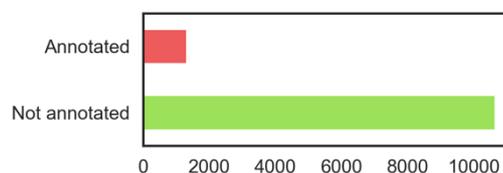


図 1 機能に関する注釈があるマイナーなスプライシングアイソフォームの数

これらの機能に関する注釈があるマイナーなスプライシングアイソフォームと、それ以外のマイナーなスプライシングアイソフォームの間では、解析した指標全てにおいて弱い傾向の違いが見られたものの (図 2)、強く相関する指標は見られなかった。このことから、機能的なスプライシングアイソフォームを推定するには特定の指標のみを用いるのではなく、複数の指標の組み合わせが必要であることが確認された。

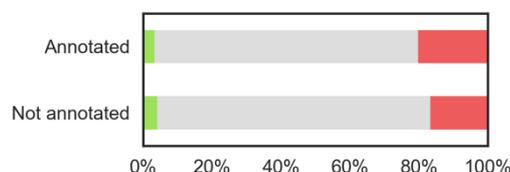


図 2 解析した指標の例。緑は機能ドメインが増えるもの、赤は機能ドメインが減少するものを示す。機能に関する注釈がある方が、機能ドメインの数が減少する傾向にある。

また PU アルゴリズムを用いた機械学習により得られたモデルを、タンパク質としての発現が確認されていない約 49,000 個のヒトマイナーなスプライシングアイソフォームに適用したところ、約 70%において機能性があると推定された。このことは、近年議論されている (Tress et al., 2017) よりも多くのスプライシングアイソフォームが、実際にプロテオームの多様性に寄与していることを示唆している。

これらの各スプライシングアイソフォームについて解析した特徴については、研究代表者がこれまでに開発、公開している選択的スプライシングの Web データベースである AS-ALPS に反映させた。

## (2) 立体構造情報に基づく低分子結合部位及び結合低分子予測法の開発

これまでに開発してきた ProLMS において、統計情報を得るのに用いるタンパク質立体構造情報を更新することで、132 種類の低分子について結合部位を予測できるようにした。さらに、ProLMS を機械学習により拡張することで、様々な酵素において重要な補欠分子族であるピリドキサルリン酸 (PLP) の結合残基予測において、大幅な改善が見られた (図 3)。

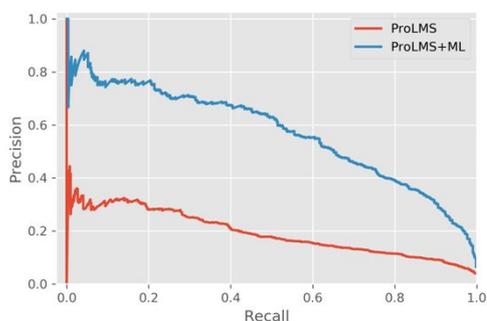


図3 PLP 結合残基予測の精度-再現度プロット。オリジナルの ProLMS (赤) に比べて機械学習による拡張後の ProLMS (青) では精度が大幅に上昇している。

また、標的タンパク質に PLP が結合するかどうかを予測する場合においても予測精度を示す指標の一つである Matthews 相関係数が 0.46 となり、この結合能予測の方針が、有効である可能性が示された。今後は、より多くの低分子に拡張し、この結合能予測法の有効性を示す必要がある。

また、ProLMS の改良の過程で得られた低分子に対する結合傾向値を、研究代表者が維持してきた低分子化合物の Web データベースである Het-PDB Navi. に反映させ、さらにユーザーインターフェースの大幅な改良を施して Het-PDB Navi2 として新たに公開した。Het-PDB Navi2 では、タンパク質との複合体で立体構造が決められている低分子について、タンパク質がどのアミノ酸残基を使ってどのように低分子を結合しているかを観察できるようにした。さらに、化学構造が類似した低分子の検索や、タンパク質の立体構造データと低分子を指定すると、ProLMS を実行してその低分子が結合する部位の予測もできるようにした。

一方で、ProLMS の拡張により結合する低分子を予測するだけでは、低分子がどのような配向で結合するかを予測することができない。そこで、低分子結合に使われるアミノ酸残基の側鎖をベクトルで表現したものをデータベース化し、低分子結合を予測したい標的タンパク質の表面でデータベースと合致するベクトルを検索することでタンパク質と低分子をドッキングする方法を新たに開発した。この手法は、低分子の結合の配向を求められるだけでなく、標的タンパク質にデータベースと合致するベクトルが存在するかどうかを指標とすることで、低分子の結合能も予測できる。そこで、112 種類の低分子について、692 個の標的タンパク質への結合低分子予測の性能を評価したところ、既存の低分子ドッキング法 (AutoDock Vina) よりも高い性能で結合する低分子を判定できることがわかった (図 4)。

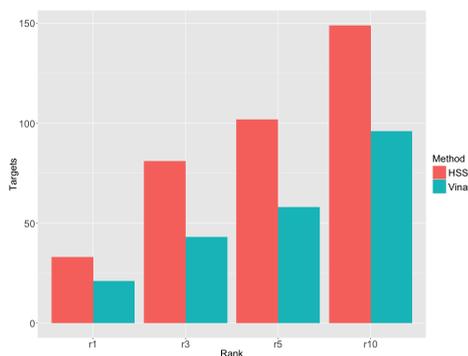


図4 結合低分子予測性能の評価 横軸が正しいリガンドの予測ランク (左から 1 位、3 位、5 位、10 位)、縦軸が標的タンパク質の数を表す。オレンジが本手法の結果、水色が AutoDock Vina の結果を表す。

以上のリガンドとタンパク質のドッキングについての研究から得られた知見を活用して、新規プロテアソーム阻害剤やカルサイトのドッキングも行った。

### (3) シングルヌクレオチドバリエーション (SNV) に基づく機能部位の解析

構築した健常人全エクソンバリエーションデータ解析システムを用いて、各 SNV の立体構造および機能部位における特徴についての解析を行なった。その結果、SNV と機能部位の間には、機能部位予測に使えるような明確な関連性が見られないことがわかった。その一方で、顕性の遺伝様式を持つ疾患において、発病メカニズムの種類が、原因となる SNV の立体構造上の位置、特に相互作用する分子の種類と強く関連していることがわかった。

さらに、疾患に関わる SNV では、スプライシングアイソフォームで共通に使われているエクソンと、選択的に使われるエクソンの間で分布の偏りがあることが示された。この解析を進展させることで、病気に関わる SNV の分布から、スプライシングアイソフォームの機能性推定を行う研究へと展開できると考えられた。

これらの解析に用いた健常人全エクソンバリエーションデータ、および、ヒトの mRNA およびタンパク質へのマッピング結果は、研究分担者が開発している Web データベースである Mutation@A Glance から検索できるようにした。

## 5. 主な発表論文等

〔雑誌論文〕(計 10 件)

S. Nakae, M. Shionyu, T. Ogawa and T. Shirai, Structures of jacalin-related lectin PPL3 regulating pearl shell biomineralization., *Proteins*, 査読有, 86, 2018, 644-653, DOI: 10.1002/prot.25491

M. Tanaka, Y. Zhu, M. Shionyu, N. Ota, N. Shibata, C. Watanabe, A. Mizusawa, R. Sasaki, T. Mizukami, I. Shiina and M. Hasegawa, Ridaifen-F conjugated with cell-penetrating peptides inhibits intracellular proteasome activities and induces drug-resistant cell death., Eur J Med Chem, 査読有, 146, 2018, 636-650, DOI: 10.1016/j.ejmech.2018.01.045

A. Hijikata, T. Tsuji, M. Shionyu and T. Shirai, Decoding disease-causing mechanisms of missense mutations from supramolecular structures., Sci Rep, 査読有, 7, 2017, 8541, DOI: 10.1038/s41598-017-08902-1

C. Shionyu-Mitsuyama, A. Hijikata, T. Tsuji, and T. Shirai, Classification of ligand molecules in PDB with graph match-based structural superposition., J Struct Funct Genomics, 査読有, 17, 2016, 135-146, DOI: 10.1007/s10969-016-9209-x

A. Hijikata, K. Yura, O. Ohara, M. Go, Structural and functional analyses of Barth syndrome-causing mutations and alternative splicing in the tafazzin acyltransferase domain., Meta Gene, 査読有, 4, 2015, 92-106, DOI: 10.1016/j.mgene.2015.04.001

[学会発表](計18件)

土方敦司、白井剛

Mutation@A Glance: ヒトにおける意義不明バリエーションの解明を目指した統合解析ツール、ConBio2017、2017年

塩生真史、郷 通子

Het-PDB Navi2: タンパク質 - 低分子複合体構造から得られる相互作用部位のデータベース、トーゴーの日2017、2017年

塩生真史、Pramote Teerasetmanakul Estimating functionality of expression-unconfirmed splicing isoforms at the protein level、第55回日本生物物理学会年会、2017年

Pramote Teerasetmanakul、塩生真史 機械学習に基づく機能未知スプライシングアイソフォームの機能性推定、第39回日本分子生物学会年会、2016年

Atsushi Hijikata, Masafumi Shionyu, Tsuyoshi Shirai

ベクトル表現化したアミノ酸残基のマッチングによるタンパク質-リガンド結合予測、第54回日本生物物理学会年会、2016年

Masafumi Shionyu

Function prediction of uncharacterized splicing isoforms、第53回日本生物物理学会年会、2015年

[その他]

ホームページ等

AS-ALPS データベース

<http://as-alps.nagahama-i-bio.ac.jp/>

Het-PDB Navi2 データベース

<http://hetpdbnavi.nagahama-i-bio.ac.jp/>

Mutation@A Glance データベース

<http://mutation.nagahama-i-bio.ac.jp>

## 6. 研究組織

### (1) 研究代表者

塩生 真史 (SHIONYU, Masafumi)

長浜バイオ大学・バイオサイエンス学部・准教授

研究者番号: 30345847

### (2) 研究分担者

土方 敦司 (HIJIKATA, Atsushi)

長浜バイオ大学・バイオサイエンス学部・プロジェクト特任講師

研究者番号: 80415273

### (4) 研究協力者

TEERASETMANAKUL, Pramote

中田 圭祐 (NAKATA, Keisuke)