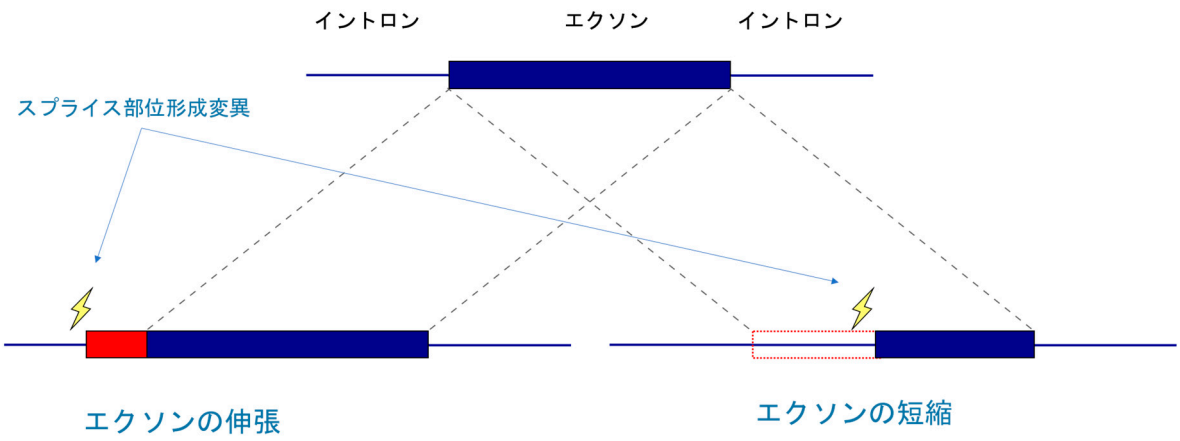
85　遺伝性疾患の原因となる偽エクソン活性化の網羅的探索	須山　幹太

【目的】遺伝性疾患の原因変異として、これまで主にコード領域内でアミノ酸置換やフレームシフトを起こすものやスプライス部位を壊すものが同定されてきた。しかし、近年、新たなスプライス部位を形成する変異により通常用いられるエクソンとは異なるエクソン（偽エクソン）が活性化されることで有害な転写産物がもたらされるために疾患となる事例が報告されるようになった。このような変異は、機能を持っていない領域で起こる変異であることから、これまで見過ごされることが多かった。そこで本研究は、スプライス部位を形成することで遺伝性の希少疾患の原因となるレアバリアントの網羅的な同定を目的とした。本研究は、遺伝性疾患について、これまであまり注目されてこなかったスプライス部位形成変異に着目することで新たな原因変異を同定し、確定診断が困難なことが多いそれらの疾患の診断率の向上に貢献することを目指すものである。

【方法】対象とする疾患は特定のものに絞らず、遺伝性疾患全般について原因変異を探索する疾患横断的なアプローチをとった。方法としては、バイオインフォマティクス解析を中心とし、公共データベースに蓄積しているヒトの変異データに対し、スプライス部位の予測プログラムを適用することで、スプライス部位形成能を持った有害な変異を探索した。得られた候補変異については、その変異を有する個人のトランスクリプトームデータを解析し、実際にその変異がスプライスに変化をもたらしているかを検証した。これらを通し、様々な遺伝性の希少疾患について、その原因となる新たな変異を同定した。

【結果】遺伝性疾患の原因となることが知られている約 4,000 遺伝子を対象に、ヒトの変異データベースに登録されている約 7 万人分の変異データの中から、既知エクソン内、あるいはその周辺に存在する変異を抽出した。それらの変異について、統計的手法および人工知能を応用したスプライス部位予測法を適用することで、各変異のスプライス部位形成能を評価した。その結果、5,656 個のスプライス部位形成変異を得た。さらに、これらの変異について、コドン読み枠のズレや終止コドンの挿入、タンパク質ドメイン構造内の挿入・欠失などにより、病原性が考えられるかを評価したところ、3,942 個の変異を病原性があると判断された。実際の変異の機能について、その変異を持つ個人の RNA-seq データで検証したところ、実際にスプライス部位を形成する変異であることが確認された。今回の手法を、原因変異が未同定の繊毛病患者から得られた変異データに適用したところ、新規原因候補となるスプライス部位形成変異の同定に成功した。

本研究で解析対象とした変異とそれらにより生成される異常なエクソン



85. 遺伝性疾患の原因となる偽エクソン活性化の網羅的探索

須山　幹太

九州大学 生体防御医学研究所 情報生物学分野

Key words：遺伝性疾患，スプライス部位形成変異，バイオインフォマティクス，疾患ゲノム解析，RNA-seq

緒　言

遺伝性疾患は、知られているだけでも約 7,000 種ある。その多くは患者数が少ない希少疾患であり、いわゆる難病と呼ばれるものである。個々の疾患の患者数は少ないが、疾患の種類が多いため、全世界では 3 億 5 千万もの人が何らかの遺伝性疾患に罹患していると言われている。遺伝性疾患の原因として、これまで多数の原因変異が同定されてきたが、遺伝性疾患サンプルのゲノムを調べても原因変異が見つかる確率は 50%ほどしかなく、まだまだ未知の原因変異が存在していると考えられる。これまで遺伝性疾患の原因変異として、主にコード領域内やスプライス部位に同定されてきた。その理由は、これらの領域は機能的に重要な領域であり、そのような領域での変異はその機能にとって有害であると容易に判断できるからである。それ以外の領域は明らかな機能的に役割を担っていないため、そこでの変異は疾患の原因としてはあまり注目されてこなかった。

近年になり、遺伝性疾患の原因変異として、機能的に重要ではないゲノム領域における変異が報告されるようになった［1］。これらの変異は機能的に重要でない領域に新たなスプライス部位を形成し、本来のエクソンとは異なる偽エクソンを取り込んだ異常な転写産物を生成することで疾患を引き起こしていることが示されている。このようなスプライス部位形成変異は、疾患原因において散発的な報告にとどまっていたが、最近私たちのグループは健常者においても一定の頻度で見られるものであることを示した［2］。しかし、多種多様な遺伝性疾患全般において、どの程度原因となっているかは不明であった。

そこで本研究は、スプライス部位を形成することで遺伝性の希少疾患の原因となるレアバリアントに注目し、遺伝性疾患全般について疾患横断的なアプローチにより原因変異の網羅的な同定を目指した。その際、疾患ゲノムデータを直接用いるのではなく、現在大規模ゲノム解析によって蓄積が進んでいる個人ゲノムデータを用いた。これらのゲノムは基本的に健常者由来のものであるが、健常者であっても遺伝性疾患の保因者であると考えられるため、それらの個人ゲノムの中に新規原因候補となるスプライス部位形成変異の網羅的な探索を行った［3］。

方　法

1. ゲノム変異データ

Genome Aggregation Database（gnomAD release 3.0）から、VCF フォーマットにて 71,702 人分のゲノム変異データを取得した。その中で、アリル頻度が 0.01 以下である約 5 億 6 千万の SNV を解析対象とした。

2. スプライス部位形成変異の予測

ある変異がスプライス部位を形成するものかどうかは、まず MaxEntScan プログラムにより評価した。MaxEntScant のスコアが正であったものについて、さらに SpliceAI プログラムにより評価した。SpliceAI のスコアが 0.8 以上のものをスプライス部位形成変異とした。

3. 個人の RNA-seq データを用いたスプライス部位形成変異の検証

スプライス部位形成変異として予測された変異について、その変異を有する個人の RNA-seq データを解析することで実際にその変異がスプライス部位形成するものであるかを検証した。この検証には GTEx データベースに登録され

ている 670 人分の個人ゲノムデータと、それらの個人から得られた全血の RNA-seq データを用いた。

スプライス部位形成変異による新規スプライス部位を精度よく検出するには、個人の RNA-seq データをその個人のゲノムに正確にマップする必要がある。そのため、まず個人ゲノムデータから得られた変異情報を反映させた参照ゲノム配列を構築し、そのゲノム配列に対して HISAT2 プログラムによりその個人の RNA-seq データをマップした。

4. 繊毛病患者のエクソームデータを用いたスプライス部位形成変異の探索

繊毛病患者 1,773 人とその非発症家族のエクソームデータを dbGaP から取得した。これらデータに見られる変異のうち、434 個の繊毛病関連遺伝子に見られるスプライス部位形成変異を探索した。

結　果

1. 遺伝性疾患の原因遺伝子に見られるスプライス部位形成変異の同定

遺伝性疾患の原因となることが知られている 4,054 遺伝子を対象に、gnomAD に登録されているアリル頻度 0.01 以下の変異について、スプライス部位形成能を評価した。スプライス部位形成変異は、既知のエクソン内あるいはエクソン - イントロン境界から 50 塩基以内に存在するものとした。その結果、5,656 個のスプライス部位形成変異を得た。これらのうち 3,942 個は、コード配列に未成熟終止コドンを挿入する、あるいは、タンパク質ドメイン構造内にアミノ酸の挿入・欠失を導入する、といった有害であると考えられる変異であった。

2. 同定したスプライス部位形成変異および RNA-seq データによる検証の例

有害であると考えられるスプライス部位形成変異と RNA-seq データによる検証の例を 2 つ挙げる。

一つ目の例は小人症 MOPD2 の原因遺伝子として知られる *PCNT* 遺伝子のイントロン中に見つかったスプライス部位形成変異である。GTEx データベースには、この変異をヘテロ接合で持つ個人が一人いた。その個人の RNA-seq から、この変異が実際にスプライス部位形成変異として機能していることが確認できた（図 1a）。この変異を持たない個人では、そのすべてにおいてエクソンの伸張が起きていないことからも、この変異がスプライス部位形成変異であるといえる。この変異により伸張した部分に終止コドンが存在するため、通常よりも短い有害なタンパク質ができると考えられる。この変異は、まだ MOPD2 の原因であるとの報告はなく、今回の結果は新規原因候補変異であることを示すものである。

二つ目の例はリソソームにグリコーゲンが異常に蓄積するポンペ病の原因遺伝子として知られる酸性 α グルコシダーゼ（*GAA*）遺伝子のエクソン中に見つかったスプライス部位形成変異である。GTEx データベースには、この変異をヘテロ接合で持つ個人が一人いた。その個人の RNA-seq から、この変異が実際にスプライス部位形成変異として機能していることが確認できた（図 1b）。この変異を持たない個人では、そのすべてにおいてエクソンの伸張が起きていないことからも、この変異がスプライス部位形成変異であるといえる。興味深いことに、この変異はシノニマス変異であるため、通常は有害変異と判断されないものであった。実際は、このエクソンの短縮によりフレームシフトが起き、下流に未成熟終止コドンを生成する。この変異は病原性変異としては未だ報告されていないものであるが、この変異の近傍に病原性変異があることからも、新たな病原性変異であることが強く示唆される。

3. 同定したスプライス部位形成変異の位置と変異スペクトラム解析

次に、同定した 5,656 個のスプライス部位形成変異について、その構造的特徴を解析した。5,656 個の変異のうち、2,218 個（39.2%）は 5'スプライス部位に、3,438 個（60.8%）は 3'スプライス部位に存在した。また、既知のイントロン中に存在するものは 2,937 個（51.9%）であり、既知のエクソン中に存在するものは 2,719 個（48.1%）であった。5'スプライス部位および 3'スプライス部位に存在するスプライス部位形成変異について、その位置と変異スペクトラム解析を行った（図 2）。5'スプライス部位では-2 位と-3 位に変異が多くみられた（図 2a）。この部位での変異により、5'スプライス部位によくみられる GYNGYN モチーフが形成されると考えられる。イントロンの末端部では変異が少なかった。それ以外の領域では、変異は一様に分布していた。3'スプライス部位では-1 位に変異が多くみられた（図 2b）。-4 位と-5 位に見られる変異は、3'スプライス部位によくみられる NAGNAG モチーフを形成するものと考えられる。5'スプライス部位と異なり、3'スプライス部位ではエクソン中の変異が少なかった。

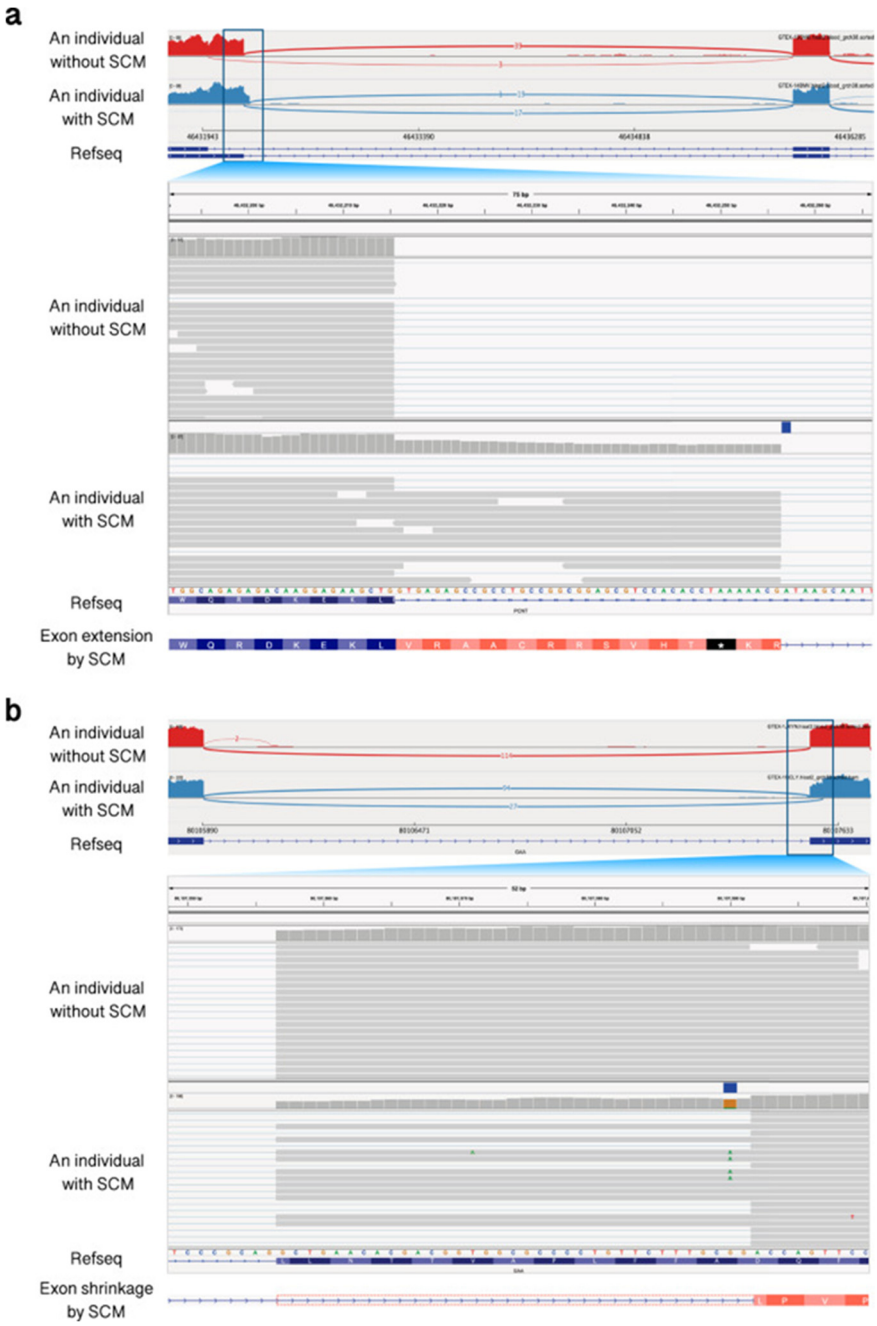


図 1. スプライス部位形成変異によるエクソンの伸張と収縮の例

a) *PCNT* 遺伝子に見られたスプライス部位形成変異によるエクソンの伸張。上段はスプライス部位形成変異を持たない個人と持つ個人の RNA-seq リードの集積およびジャンクションリードを表す Sashimi プロットを示す。下段は、スプライス部位形成変異近傍を拡大したもので、実際のリード配列を示す。エクソンの伸張により変化したアミノ酸配列をその下に示す。

b) *GAA* 遺伝子に見られたスプライス部位形成変異によるエクソンの収縮。上段はスプライス部位形成変異を持たない個人と持つ個人の RNA-seq リードの集積およびジャンクションリードを表す Sashimi プロットを示す。下段は、スプライス部位形成変異近傍を拡大したもので、実際のリード配列を示す。エクソンの収縮により変化したアミノ酸配列をその下に示す。

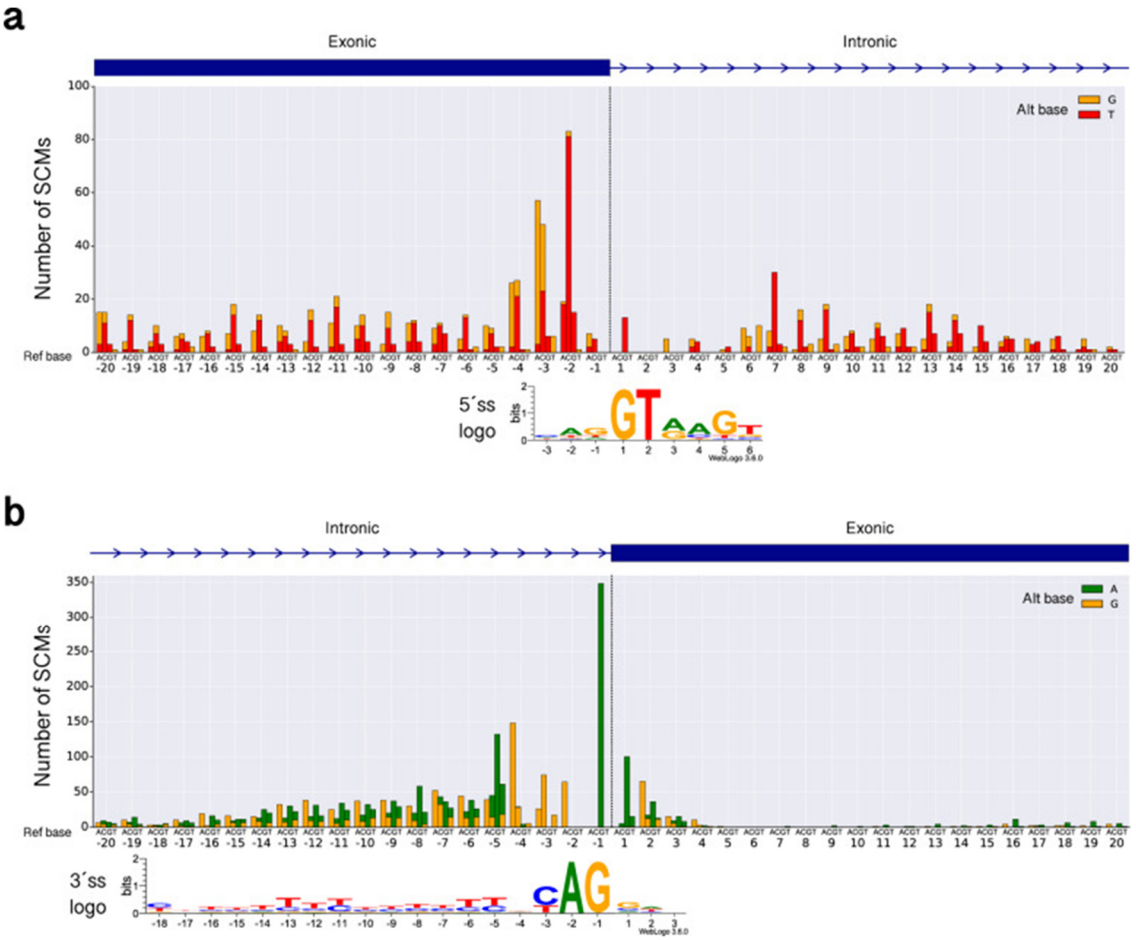


図 2. スプライス部位形成変異の位置と変異スペクトラム

a) 5'スプライス部位におけるスプライス部位形成変異の位置と変異スペクトラム。

b) 3'スプライス部位におけるスプライス部位形成変異の位置と変異スペクトラム。

それぞれのプロットの下にスプライス部位のモチーフを WebLogo で示す。

4. 繊毛病患者のエクソームデータ解析により同定した新規原因候補変異

以上によりスプライス部位形成変異が多くの遺伝性疾患の原因候補変異となることが示されたので、次に、原因変異が未知の疾患エクソームデータからの新規原因変異の探索を試みた。ここでは繊毛病患者のエクソームデータを解析対象とした。その結果、新たに 38 個の変異がスプライス部位形成変異として繊毛病の原因変異となっている可能性を示すことに成功した。そのうちの一つは繊毛病の原因遺伝子として知られている *CSPP1* 遺伝子のイントロン中に見つかったスプライス部位形成変異である（図 3）。この変異は、独立した家系の 3 人の患者においてホモ接合で見られ、また、その非発症家族 1 名にヘテロ接合で見られた。このことから原因変異であることが強く示唆されるものである。この変異によりエクソンの伸張が起き、その伸張部分に未成熟終止コドンを生成することがわかった。このすぐ下流には既知の病原変異が存在することからも、今回見つけたスプライス部位形成変異は繊毛病の新規原因変異であると考えられる。

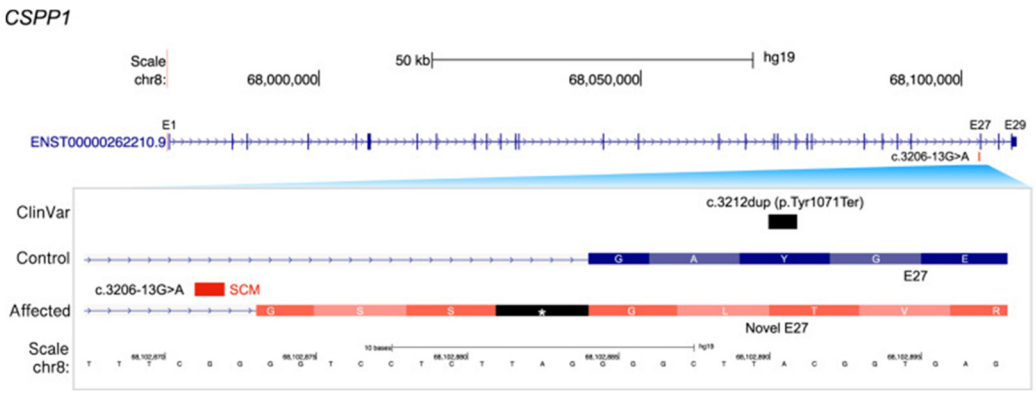


図 3. 繊毛病の原因遺伝子である *CSPP1* 遺伝子に見られたスプライス部位形成変異
上段に *CSPP1* 遺伝子の構造を示す。下段にスプライス部位形成変異の近傍を拡大し、エクソンの伸張により変化したアミノ酸配列を示す。

考　察

本研究では、遺伝性疾患の原因になることが知られている 4,054 遺伝子の中に病原性があると考えられる 3,942 個のスプライス部位形成変異を同定することに成功した。この数は、遺伝性疾患における既知原因変異数の 8.5%に相当し、疾患原因変異探索においてスプライス部位形成変異も考慮することの重要性を示すものである。今後、これらの変異が疾患発症への関与が実験的にも検証されれば、遺伝性疾患の診断への応用が可能となる。また、この成果は、疾患原因となっている新たに形成されたスプライス部位を抑制するような核酸医薬品の開発による、遺伝性疾患の治療の可能性にもつながるものである。

共同研究者

本研究の共同研究者は、九州大学生体防御医学研究所情報生物学分野の坂口愛美である。

文　献

1) Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet. 2017 Sep;136(9):1093-1111. Epub 2017 May 12. PMID 28497172 DOI: 10.1007/s00439-017-1809-4

2) Sakaguchi N, Suyama M. In silico identification of pseudo-exon activation events in personal genome and transcriptome data. RNA Biol. 2021 Mar;18(3):382-390. Epub 2020 Aug 30. PMID: 32865117 DOI: 10.1080/15476286.2020.1809195

3) Sakaguchi N, Suyama M. Pervasive occurrence of splice-site-creating mutations and their possible involvement in genetic disorders. NPJ Genom Med. 2022 Mar 18;7(1):22. PMID: 35304488 DOI: 10.1038/s41525-022-00294-0