

## = 総 説 =

## 網羅的遺伝子診断としての次世代シーケンス結果の評価

山 本 俊 至

**要旨** 次世代シーケンスによる網羅的遺伝子解析により、多くの疾患関連遺伝子が明らかにされてきた。このことにより、網羅的遺伝子解析は研究レベルから臨床診断への応用に利用目的が広がってきている。次世代シーケンスを用いた解析においては、実験室レベルで検体を調整する過程を経てシーケンサーにかけられ、自動的に解析結果がデジタルファイルとして得られるが、最終的には疾患について専門的な知識を持つ医師によって判断されなければ疾患関連バリエーションに辿り着くことは難しい。その判断過程においては、バリエーションの意味づけや頻度情報の扱い方、様々なデータベースの利用方法、トリオ解析の重要性を認識しておく必要がある。その上で、filtering や curation がどのように行われるべきか、さらに ACMG ガイドラインに沿ってどのように判断すべきかについて概説した。

**見出し語** 次世代シーケンス、網羅的遺伝子解析、バリエーション、データベース、トリオ解析

## はじめに

次世代シーケンスによる網羅的遺伝子解析により、これまで知られていなかった多くの疾患関連遺伝子が明らかになってきた。特に小児神経疾患領域においては、てんかん関連遺伝子や知的障害、発達障害などに関連する遺伝子が数えきれないほど明らかになってきたことは小児神経科医の多くが知るところであり<sup>1)</sup>、筆者も多くの診断に関わってきた<sup>2)</sup>。特に2015年からは、日本医療研究開発機構 (Japan Agency for Medical Research and Development; AMED) による未診断疾患イニシアチブ (Initiative on Rare and Undiagnosed Diseases; IRUD) によって未診断疾患患者の診断が強力に推し進められており、次世代シーケンス (next-generation sequencing; NGS) による網羅的遺伝子解析はより身近な診断方法となってきているため、自らが診療している未診断難病患者の診断が次々と明らかになる様をまざまざと体験している小児神経科医も少なくないと思われる<sup>3)4)</sup>。

一方、IRUD などでは最終的な診断だけが主治医に返却され、NGS 結果がどのように評価されているのかブラックボッ

クスのままであり、中にはその解析過程に興味を持っている医師も多くいることと思われる。それが証拠に、日本小児神経学会における実践教育セミナーなどで NGS データのハンズオンセミナーを開催すると、すぐに定員がいっぱいになる状態が続いている。興味はあるけど実際に体験する機会に恵まれないということを示唆している。そのようなセミナーで講師を務めると、「このセミナーの内容を記載したテキストはないのか」としばしば質問されるが、残念ながら「そのようなテキストはありません」と答えるしかない。というのも、技術の進歩は日進月歩というより秒進分歩の状況であり、テキストを著している間にその内容はたちまち陳腐で古く、使い物にならないものになってしまうからである。とは言え、今現在の NGS データの解析方法について、著しておくことは意味がなくはないだろう。

本稿の目的は、今現在の NGS データ、特にアノテーション (annotation) 後の filtering と curation について、初学者向けに解説することにある。

## I NGS 解析の実際

IRUD で主に行われている NGS 技術はほとんどの場合、全エクソーム解析 (whole exome sequence; WES) である。ヒトゲノムは全部で30億塩基対からなるが、遺伝子をコードしている重要な部分は、このうちわずか1~2%を占めるエクソン領域である。そのため、この領域だけを解析するのが効率的とされている。

現在主流となっている NGS の機種は、そのほとんどで short read sequence 技術が用いられている。最大で125塩基程度の短い配列を超大量に読み、ヒトのリファレンス配列上に

東京女子医科大学大学院医学研究科先端生命医科学専攻遺伝子医学分野

東京女子医科大学遺伝子医療センターゲノム診療科

連絡先 〒162-8666 東京都新宿区河田町 8-1

東京女子医科大学遺伝子医療センターゲノム診療科  
(山本俊至)

E-mail: yamamoto.toshiyuki@twmu.ac.jp

(受付日: 2020. 2. 25, 受理日: 2020. 2. 25)

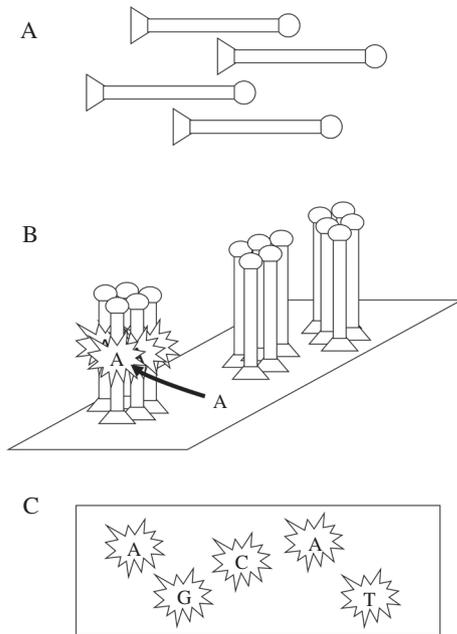


図1 次世代シーケンスの原理

A: 短く断片化した DNA の両端にアダプターを付けてライブラリー化する。B: フローセルのガラス表面に付着させてクラスター形成させ A, C, G, T の4つの塩基を1ずつ取り込ませる。C: フローセルを真上から撮像し、フローセルに取り込まれた蛍光色素の変化をとらえ、A, C, G, T の変化として認識する。

屋根瓦状にマップさせ、重ね合わせることで、ヘテロ接合バリエーションなどを検出することが可能となる<sup>5)6)</sup>。

実際の実験工程としては、患者サンプル DNA はまず最初に超音波などにより短く断片化され、両端にアダプターを結合させてライブラリーとし、それをフローセルというガラス基板の上にランダムに貼り付けることで準備が完了する(図1A, B)。フローセルを次世代シーケンサーに装着したあとは、ライブラリーに A, C, G, T の4つの塩基が順に取り込まれていくが、それぞれの塩基は4種類の蛍光色素で標識されており、1塩基取り込まれるごとにフローセルを真上から画像として取り込んでいくことができる(図1C)。最終的に取り込まれた蛍光色素の順番を塩基の種類に置き換えていくことにより、ライブラリーの塩基配列として認識できるようになる。フローセル上には何万個ものライブラリーが付着しているため、ライブラリーの数だけのシーケンス配列が得られることになるが、1つのライブラリーから得られた125塩基の情報のことを read と呼ぶ。

ここまでで得られたデータは fastq というファイルとして保存されている。次にこのファイルをヒトリファレンス配列に annotation しなければならない。これによって、ヒトリファレンス配列と相違のある塩基の情報が、Bam や vcf といったファイルとして展開される。これらの作業には複数の解析ソフトをパイプラインとして用いる必要があり、比較的スペックの高いコンピューターやサーバー、Linux OS などを必要と

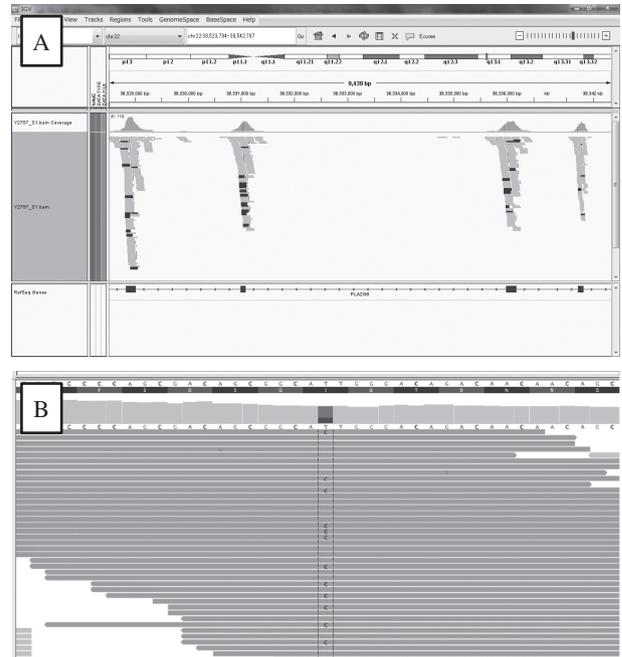


図2 IGV での参照画面

A: 多くの read がエクソン領域にマップされていることがわかる。B: 拡大表示すると、塩基が一致しない場所はエラー表示されていることがわかる。Integrative Genomics Viewer 使用

(IGV : <https://software.broadinstitute.org/software/igv/>)

する必要があるが、ベンダーが提供する VariantStudio (Illumina) や SureCall (Agilent Technologies) と言った専用ソフトを用いれば、個人のノートパソコンでも解析可能である。ただ、学会主催のハンズオンセミナーなどでデータを取り扱う場合、専用ソフトを使用することが難しいため、これらのソフトを用いて得られたデータをさらに csv ファイルとして export し、excel の機能を用いて解析するようにしている。

## II アノテーション (annotation)

NGS データの評価方法を解説する前に、実際に NGS を行って得られたデータはどのようなものなのか解説しておきたい。

先に述べたように、次世代シーケンサーではせいぜい125塩基対程度の短い DNA 断片の配列が超大量に得られる。Annotation によって1つ1つの read がヒトリファレンス配列に張り付けられる (mapping) が、WES がターゲットとしているエクソン領域には大量の read が瓦状に張り付いており、その状態は Integrative Genomics Viewer (IGV : <https://software.broadinstitute.org/software/igv/>) という無償ソフトを用いると可視化して確認することができる(図2A)。IGV でデータを可視化するためには Bam, Bam.bai, vcf という3種類のファイルが必要となる。IGV で視覚的に確認できるように、当該部分にどれだけたくさんの read が張り付いているかは depth と表現される。次世代シーケンサーの1つ1つの塩基の違いにはエラーも結構紛れているため、depth が深いほどその領

①ゲノムの位置を入力

②Referenceのバージョンを確認

③MLC1遺伝子領域であることを確認

④ゲノム上はCではなくG

図3 UCSC genome browser の参照方法

UCSC genome browser (<https://genome.ucsc.edu/>) より

域のデータの信頼性が向上する。IGVで表示されたreadは完全にマッチしていればグレーで表示されるが、バリエーションがあるなどのためマッチしない部分はカラーで表示される(図2B)。ヘテロ接合バリエーションの場合、2本の染色体上にある2つのアレルのうちの1つだけにバリエーションがあるため、readの約半数にバリエーションが見られなければならない。もしdepthが100と十分にあり、そのうちおよそ半数の50readにバリエーションが認められるような場合、エラーの可能性は考えにくい。逆にdepthが5程度しかない場合、2つのreadで1塩基置換があったとしても、そのようなデータを評価の対象として採用すると質の高い解析結果は得られない。したがって、filteringの過程でdepthの低い領域のデータはquality controlとして排除される場合が多い。

### III NGS データ

ハンズオンセミナーではテキストの有無だけではなく、練習に使用するデータの在処にも質問が及ぶことがある。本邦では、個人情報保護法の影響で、次世代シーケンスによる多数のバリエーションがまとまった情報は、個人情報と見なされ、保護の対象となっている。したがって、不特定多数が閲覧できるweb siteなどには公開されておらず、AMEDデータベースなどでもセキュリティが掛けられており、閲覧するためには申請が必要である。

一方、海外のデータベースやonline journalなどでは論文となったデータなどが自由にダウンロードできる状態で公開されている。しかし、そのほとんどはfastqやvcfファイルの状態で開催されており、csvファイルは著者が調べた限り見つからなかった。fastqやvcfファイルの展開は初学者にはややハードルが高いと思われる。もし演習用にcsvファイルが必要という場合は、著者まで直接連絡いただきたい。

### IV バリエーションの記載方法

ヒトのゲノム配列はそれぞれの染色体の短腕末端を1として長腕末端に向かって番号が振られている。chr22:

50,518,816と記載した場合、22番染色体短腕末端から50,518,816番目の塩基を指している。UCSC genome browser (<https://genome.ucsc.edu/>)という統合データベースを用いれば、この記載を入力することによって視覚的に閲覧できる(図3)。ただし、ヒトゲノム標準配列(リファレンス)はヒトゲノムプロジェクト終了後、数年毎にバージョンアップされており、使われるバージョンによってゲノム上の位置情報が異なっており、chr22:50,518,816と記載した場合でも、それがどのバージョンによる位置情報かわからなければまったく意味がなく、chr22:50,518,816(hg19)などと記載しなければならない。UCSCで使用できるヒトゲノムリファレンス配列の最新バージョンはGRCh38/hg38であるが、様々なデータベースへの紐付けは十分ではなく、最近の研究論文においても、ひとつ前のバージョンであるGRCh37/hg19が頻用されている。

次に遺伝子の配列情報に与えるバリエーションの影響を記載する方法について解説しておきたい。ゲノム全体から位置を辿る場合は上記の位置情報が有用であるが、各バリエーションの評価を行うためには、それらがどの遺伝子のどのような変化に関わっているかなど、annotation後に付与される情報を参照することになる。特定の遺伝子におけるバリエーションを記載する方法はHuman Genome Variation Society(HGVS)によって国際的に統一されており、NM\_139202.3(MLC1):c.278C>Tというような記載方法となっている。NM\_139202.3というのはRefSeqという各遺伝子の配列データのIDである。「c」はcomplimentary DNAのCを意味しており、当該遺伝子の開始コドンであるmethionine(Met)をコードするATG配列の最初の「A」を1番とすることになっている。これによって何番目の何塩基が何塩基に置換されているかが示される。RefSeq番号は遺伝子によっては2つ以上の番号が付与されている場合がある。1つの遺伝子から複数のスプライシングバリエーション、いわゆるアイソフォームが発現されることがあるため、どのアイソフォームを指しているか規定しなければならないのである。NM\_139202.3(MLC1):c.278C>Tは別のRef-

Seq を用いて NM\_015166.3:c.278C>T と記載しても同じ意味となる。記載方法に誤りがないかどうかは Mutalyzer (<https://mutalyzer.nl/>) という web soft で確認できる。

アミノ酸が置換されるミスセンス以外では、“del”は deletion の略で当該塩基の欠失，“dup”は duplication の略で当該塩基がもう 1 回繰り返されている。“ins”は insertion の略で何らかの塩基が挿入しており、どの塩基とどの塩基の間にどのような配列が挿入されたか記載される (ex. NM\_139202.2:c.1052\_1053insCGGGGAGGTGAGTGGCCTGTGGGGTGGGGGTGC)。しばしば一般的な言葉として「indel」という用語が使われるが、これは insertion や deletion を総称したものである。一方、稀ではあるが、本当に insertion と deletion が同時に生じることもある。この場合 NM\_015166.4:c.337\_353delinsG (p.Ile113fs) などと記載される。つまり、c.337\_353 が deletion (欠失) し、そこに G が insertion (挿入) しているのである。

一方、当該塩基の変化によってアミノ酸配列にどのような影響があったかを記載するのが NM\_139202.3:p.Ser93Leu という記載方法である。これによって何番目のアミノ酸がどのように変化したかが理解できる。DNA 配列と組み合わせると同時に表示する場合には NM\_139202.3(MLC1):c.278C>T (p.Ser93Leu) と記載される場合が多い。ちなみにこの変異は日本人の *MLC1* common 変異である<sup>7)</sup>。アミノ酸の表示は 3 文字で表す場合と 1 文字で表す場合があるが、どちらも構わない。NM\_139202.2:c.973C>T は stop gain を生じるナンセンス変異であり、p.Gln325\* と記載される。

イントロン領域の変化は cDNA として記載することができない。そのため、exon-intron boundary の塩基から前、あるいは後ろにどれだけ離れているかで記載することになっており、たとえばあるエクソンのすぐ次で、イントロンの最初の場所の変化は、c.714+1 G>A のように記載する。一般的にエクソンの直後(イントロンの最初)の 2 塩基は GT であり、次のエクソンが始まる直前(イントロンの最後)の 2 塩基は AG である。この配列は共通しており、この部分の変化はほとんどの場合、そのエクソンのスキッピングなどのスプライシング異常を来す。ただ、どのようなスキッピングに繋がるかは正確に予測することが困難なため、細胞内で発現している RNA の配列を実際に明らかにできた場合以外はアミノ酸配列の変化まで記載しないことがほとんどである<sup>8)</sup>。

UCSC で閲覧すると視覚的によく理解できるが、遺伝子は 2 本鎖 DNA のどちらか一方にコードされている。したがって、遺伝子がコードされている向きは 2 種類ある。染色体の位置で言うと短腕から長腕に向かって順列になっている方をセンス、逆向きをアンチセンスという場合がある。*MLC1* はアンチセンス鎖にコードされている。そのため遺伝子変化として NM\_139202.3:c.278C>T と記載されるが、ゲノムの上では chr22:50,518,816 G>A (hg19) と表記される (図 3)。アンチセンスの場合、相補的配列になっているからである。

## V 一塩基多型

網羅的遺伝子解析では、得られたデータを単に標準ヒトゲノム配列と比較し、異なっている部分を抽出しているだけであり、実に膨大なバリエーションが抽出される。しかし、そもそも標準ヒトゲノム配列というのは、ある一人の配列をとりあえずの標準として規定しただけのことであり、それが「正常」というものではない。ヒトゲノムには多様性があり、数百塩基対毎に一塩基多型 (single nucleotide polymorphism; SNP) が存在することは以前から知られており、膨大なバリエーションが SNP データベースに蓄積されている。一般的に、population 中で頻度の低い方をマイナーバリエーションと呼び、その頻度のことを minor allele frequency (MAF ; マフ) と表現する。しかし、SNP の中には、global な MAF は非常に低いのに、日本人における MAF が 100 近い値を示すようなものもある。つまり当該バリエーションは日本人特有の SNP であり、疾患とは無関係ということになる。

見つかったバリエーションが既知のものであれば、アノテーション時に SNP ID の情報が付与される。また、gnomAD (<https://gnomad.broadinstitute.org/>) などの 1,000 人ゲノムによる一般 population における頻度情報も付与される。実際 filtering を経験すると実感できると思うが、解析対象において認められたバリエーションのうち、そのほとんどは既に明らかなバリエーションであって、一般 population 中の頻度情報まですでにわかっている。希少疾患の場合、このような既知のバリエーションを除くだけで一気に候補バリエーションにまで辿り着く可能性がある。

## VI バリデーション (validation)

さてここからは、いよいよ NGS データの validation についてである。ここでは WES 解析を行い、fastq ファイルが得られ、annotation も終了し、vcf ファイルを展開し、全てのバリエーションを csv ファイルとして export できた状態として解説したい。Windows 版の excel を想定している。

Csv ファイルは excel を用いて開くことができる。先頭の行は各カラムのアイテムが表示される。左下に行数が表示されるが、バリエーションの総数と同じ意味である。まず、先頭行や、遺伝子名が表示された先頭列がスクロールしても見えなくならないようにしたい。その場合、固定したい場所のすぐ隣の列をクリックした状態でエクセル機能の「表示>ウィンドウ枠の固定」をクリックする。次に「データ>フィルター」をクリックする。これにより先頭行にプルダウンメニューが追加され、これで準備完了である (図 4)。

まず最初に「Filters」列のプルダウンメニューを展開し、QC がパスしたものだけを残す。この場合、選択肢をいったん全て無選択状態にした上で、Pass だけを選択すると良い。次に「Read Depth」が 20 以上のものだけを残す。この場合、「指定の値以上」を選択して 20 と入力する。次に「Conse-

① ウィンドウ枠の固定      ② データ>フィルター機能を用いてプルダウンメニューを表示

③ Consequenceから機能変化を来すものだけを選択

図4 エクセルを用いた filtering の実際

quence」から「synonymous」など、変異による影響が想定されにくいものを除外する。Synonymous とはいわゆるサイレント変異のことであり、塩基置換があってもアミノ酸が変化しないもののことである。

ここまでの絞り込みによってもまだ1,000を超えるバリエーションが残っているため、疾患の遺伝形式を予測しながらどういうバリエーションが想定されるかを考えなければならない。もし de novo によるヘテロ変異を想定しているなら、一般 population 中には認められないはずである。そのため、「Allele Freq」から0、あるいは空白だけを残す。さらに希少なバリエーションであればデータベースにも登録されていないはずなので、「dbSNP ID」列が空白だけのものを残す。ここまでの操作で数十個程度にまで絞り込まれているはずである。

ここまでは比較的機械的な操作であり、いわゆる filtering と呼ばれる。あとはしらみつぶしに各バリエーション、あるいは遺伝子にあたり、当該患者の臨床症状を説明できるかどうかを丹念に調べていく必要があるが、小児神経専門医が先天性腎疾患の遺伝子の絞り込みができるかというそれはなかなか困難であり、これ以降の操作は、その疾患に精通していないとなかなか難しい面がある。ここからの検索は filtering とは別に curation という概念が当てはまる。Curator というのは博物館の学術員を指す言葉であり、絵画などの価値を見分けるには専門的な知識が必要である。それと同じで、遺伝子、あるいはバリエーションの解釈に精通していないと最後の curation はなかなか困難である。

もしトリオ解析をしていれば、予測される遺伝形式に沿った filtering も可能である。de novo を想定している場合には、両親に認められるものを除外すれば良いし、X連鎖を予測するならば、父にはなく、母とだけ共有しているものを選択すれば良い。あとに述べるが、最終的な評価において、両親サン

プルにおける確認は非常に重要であり、病的と考えられても両親解析を行っていないとエビデンスが足りないため、疾患関連と断定できないこともある。特に新規変異の場合は、両親ともに網羅的解析を行わない場合でも、該当部分の確認だけでも両親解析を行い、de novo であることを確認する作業が必要である。

## Ⅶ 統合データベース

Curation においてしばしば用いられるのがデータベースである。特に統合データベースとしての UCSC genome browser は必須となる。Genome Browser の「Position/Search Term」コラムに遺伝子名を入力することにより検索を開始できる。検索をかけ、一番上位に「UCSC Genes」のリストが表示されるので、最上位リンクをクリックする。当該遺伝子が位置する染色体が左を短腕、右を長腕として横に表示される。関心領域が赤で示されるが、その部分が下の viewer に表示されている。Viewer のさらに下には夥しい数のラジオボタンが配置されているが、これにより viewer で表示する情報を取捨選択可能である。

UCSC Genes として表示された遺伝子名をクリックすると、その遺伝子についての「Description and Page Index」のページにジャンプする。ここには RefSeq ID とともに、当該遺伝子の機能や疾患との関わりなどの情報が記載されている。さらにその下の「Sequence and Links to Tools and Databases」の table には、様々な外部のデータベースへのリンクが張られている。このうち curation で頻用するのが OMIM と PubMed である(後述)。

さらにその下の「RNA-Seq Expression Data from GTEx」という情報も非常に参考になる。ここにはヒトの体の組織内での遺伝子発現量が比較表示されている。小児神経疾患関連遺

表1 病的バリエントを支持する要素の分類 (文献9) より引用改変)

| 評価                   | 解説            | 付記   |
|----------------------|---------------|--|
| <b>Very strong</b>   |               |  |
| 病的であることが非常に強く示唆されるもの |               |  |
| PVS1                 | 機能喪失変異        | ナンセンス変異やフレームシフト変異などにより機能喪失することが明らかなもの                |
| <b>Strong</b>        |               |  |
| 病的であることが強く示唆されるもの    |               |  |
| PS1                  | 既知の変異         | まったく同じアミノ酸変異が病的として確立している                             |
| PS2                  | de novo 変異    | 両親ともに変異がない   |
| PS3                  | 機能障害確立変異      | タンパク機能が障害されることが何らかの実験的方法で証明され確立している                  |
| PS4                  | 患者で高頻度        | 一般頻度に比べ、患者における頻度が著しく高い                               |
| <b>Moderate</b>      |               |  |
| 病的であることが示唆されるもの      |               |  |
| PM1                  | 特定領域の変異       | ホットスポット、あるいは機能ドメインに位置している                            |
| PM2                  | 患者特有変異        | コントロールでは認められない                                       |
| PM3                  | 両アレルに存在       | 劣性の場合トランスに位置しており複合ヘテロである                             |
| PM4                  | インフレーム変異      | フレームシフトせずタンパクの長さのみ変化                                 |
| PM5                  | 既知ミスセンス変異の場所  | 過去に別のアミノ酸に変化する変異の報告がある                               |
| PM6                  | 推定 de novo 変異 | de novo と考えられるが両親未解析                                 |
| <b>Supporting</b>    |               |  |
| 病的であることを支持するもの       |               |  |
| PP1                  | 家系内連鎖変異       | 複数の患者がいる家族において                                       |
| PP2                  | ミスセンス変異による疾患  | 機能喪失ではなく、ミスセンス変異によることが特徴付けられ、ミスセンス変異がほとんどの場合病的である遺伝子 |
| PP3                  | ダメージ予測で病的     | 複数の予測ツールで支持されている                                     |
| PP4                  | 表現型が特徴的       | 症状から強く示唆される  |
| PP5                  | 病的とする報告がある    | 検証不能であるが信頼できる雑誌などに報告されている                            |

Richards らの論文<sup>9)</sup>の Table 3 を翻訳改変。訳文は私見による。

伝子を検索する場合には、「brain」で発現が多い遺伝子に注目する、ということも可能である。

## VIII その他のデータベース

先に述べた OMIM は、もともと Victor A. McKusick が作成し、書籍として出版されていたメンデル遺伝病のカタログを起源としており、現在は無料のオンラインカタログとして公開されている (Online Mendelian Inheritance in Man: <https://omim.org/>)。カタログ番号が 1 から始まる疾患は常染色体優性遺伝性であり、2 から始まる場合は常染色体劣性、6 から始まる場合は X 連鎖であることを示している。OMIM は新たな知見が論文として公開されたものも順次掲載されていくが、掲載まである程度時間を要するため、最新の知見は PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) で検索する必要がある。

患者の表現型に関連する候補バリエントが見つかった際、そのバリエントの意義がどのように評価されているかを ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) で検索することにより評価できる。既知変異であれば ClinVar に pathogenic か benign か、あるいは uncertain かなど、評価が示されている。ただし、uncertain と見なされていても評価が変更される可能性があるため、あくまでも参考とすべきである。

筆者は Human protein atlas (<https://www.proteinatlas.org/>) というサイトもしばしば利用する。当該遺伝子によってコードされるタンパクがどの臓器で発現しており、さらにその組織

の細胞内局在、in-situ で使用できる抗体まで調べることが可能である。

## IX ACMG ガイドラインによる評価

2015 年、アメリカ臨床遺伝学会 (American College of Medical Genetics and Genomics; ACMG) からバリエントの評価のためのガイドラインが出された<sup>9)</sup>。現在、患者におけるバリエントを論文投稿する際、ほとんどの場合このガイドラインに沿って評価しているかどうかを査読段階で問われる。全てのバリエントは病的 (pathogenic)、病的推定 (likely pathogenic)、意義不明 (uncertain significant)、良性推定 (likely benign)、良性 (benign) のいずれかに 5 つに分類されるが、患者の診断においては病的かどうかを問われるため、本稿では病的と病的推定についてだけ解説する。

評価因子は表 1<sup>9)</sup>にまとめたように、より強く働く因子の順に very strong, strong, moderate, supporting の 4 つの因子に分けられている。最も強く働く 1 つ目の very strong 因子では機能喪失 (LoF; loss-of-function) 変異であるかどうかだけが問われる。ナンセンス変異やフレームシフト変異などの LoF 変異であれば PVS1 が該当する。次の strong 因子では、もし当該変異が新規 de novo 変異であれば PS1 は当てはまらないが PS2 は該当することになる。以下同様に、どの因子が当てはまるか評価する。論文投稿の際には当てはまる因子全ての記載が求められる場合がある。

表2 病的あるいは病的推定バリエントのクライテリア (文献9)より引用改変)

|                                       |
|---------------------------------------|
| 病的バリエント (pathogenic variant)          |
| (i) 1つのPVS1かつ                         |
| (a) 1つ以上のPS1~4または                     |
| (b) 2つ以上のPM1~6または                     |
| (c) 1つのPM1~6かつ1つのPP1~5または             |
| (d) 2つ以上のPP1~5                        |
| (ii) 2つ以上のPS1~4または                    |
| (iii) 1つのPS1~4かつ                      |
| (a) 3つ以上のPM1~6または                     |
| (b) 2つのPM1~6かつ2つ以上のPP1~5または           |
| (c) 1つのPM1~6かつ4つ以上のPP1~5              |
| 病的推定バリエント (likely pathogenic variant) |
| (i) 1つのPVS1かつ1つのPM1~6または              |
| (ii) 1つのPS1~4かつ1つか2つのPM1~6または         |
| (iii) 1つのPS1~4かつ2つ以上のPP1~5または         |
| (iv) 3つ以上のPM1~6または                    |
| (v) 2つのPM1~6かつ2つ以上のPP1~5または           |
| (vi) 1つのPM1~6かつ4つ以上のPP1~5             |

Richardsらの論文9)のTable 5を翻訳改変。訳文は私見による。

全ての評価を行ったら、表2<sup>9)</sup>に沿って病的、あるいは病的推定に該当するかどうか判断する。見つかった変異がde novoのナンセンス変異であれば(i-a)に分類されるので、病的であると言える。一方、状況的にde novoが疑われても、両親の解析が行われていない場合、PS2ではなくPM6となり、このため病的であるとは評価できなくなってしまうことがあるので注意が必要である。

表の内容は引用文献中の図を翻訳改変したものであり、用語も私見に基づく。実際に評価を行う際には原著を参照されたい。

### おわりに

網羅的な遺伝子解析を行う場合のバリエントの評価方法に

## Evaluation and diagnosis of next-generation sequencing results as comprehensive genomic analyses

Toshiyuki Yamamoto

*Institute of Medical Genetics, Tokyo Women's Medical University, Tokyo*

Comprehensive gene analysis using next-generation sequencing has revealed many disease-related genes. As a result, the purpose of using comprehensive gene analysis has been expanding from the research level to the application to clinical diagnosis. In the analysis using next-generation sequencing, samples are prepared at the laboratory level, processed by a sequencer, and the analysis results are automatically obtained as a digital file. It is difficult to reach disease-related variants unless filtering and curation by expert physicians. In the process, it is necessary to recognize the meaning of variants, how to handle frequency information, how to use various databases, and the importance of trio analysis. Here, it is described how filtering and curation should be done, and how to make decisions in accordance with the ACMG guidelines.

*No To Hattatsu 2020;52:361-7*

ついて概説した。本稿で扱った内容は日々進歩しているため、あくまでも現時点でのものである。実際に評価する際には、常に最新の情報を参照することが重要である。

著者の利益相反：本論文発表内容に関連して開示すべき事項なし。

### 文 献

- 1) 山本俊至, 監修, 日本小児神経学会, 編. 症例でわかる小児神経疾患の遺伝学的アプローチ. 東京: 診断と治療社, 2019.
- 2) Yamamoto T, Imaizumi T, Yamamoto-Shimajima K, et al. Genomic backgrounds of Japanese patients with undiagnosed neurodevelopmental disorders. *Brain Dev* 2019;**41**:776-82.
- 3) Adachi T, Kawamura K, Furusawa Y, et al. Japan's initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. *Eur J Hum Genet* 2017;**25**:1025-8.
- 4) Adachi T, Imanishi N, Ogawa Y, et al. Survey on patients with undiagnosed diseases in Japan: potential patient numbers benefiting from Japan's initiative on rare and undiagnosed diseases (IRUD). *Orphanet J Rare Dis* 2018;**13**:208.
- 5) 菅野純夫, 福嶋義光, 監訳. ゲノム医学. 東京: メディカル・サイエンス・インターナショナル, 2016.
- 6) 山本俊至. 次世代シーケンサーによる遺伝子解析. *小児科* 2011;**52**:1591-7.
- 7) Shimada S, Shimajima K, Masuda T, et al. MLC1 mutations in Japanese patients with megalencephalic leukoencephalopathy with subcortical cysts. *Hum Genome Var* 2014;**1**:14019.
- 8) Miyamoto S, Nakashima M, Ohashi T, et al. A case of de novo splice site variant in SLC35A2 showing developmental delays, spastic paraplegia, and delayed myelination. *Mol Genet Genomic Med* 2019;**7**:e814.
- 9) Richards S, Aziz N, Bale S, et al; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;**17**:405-24.