

## 論文の内容の要旨

論文題目 エキソーム解析データを用いた構造変異検出パイプラインの構築と神経疾患の診断への応用

氏名 田中 真生

次世代シーケンサーを用いた遺伝子解析法は、近年の解析技術の進歩およびコストの削減に伴い、メンデル遺伝形式の遺伝性疾患の診断ツールとして、急速に使用されるようになってきている。しかし、エキソーム解析では、エクソン領域内の一塩基置換変異および短い挿入・欠失変異の検出は可能であるが、構造変異については通常の解析方法では検出ができない。構造変異は通常 1 kb 以上の塩基配列が挿入、欠失あるいは重複することにより生じる遺伝子変異を指すが、神経内科領域においても構造変異が原因で生じる疾患は少なくない。*PMP22* 遺伝子の重複によるシャルコー・マリー・トゥース病 (CMT1A)、*PMP22* 遺伝子の欠失による遺伝性圧脆弱性ニューロパチー (HNPP)、*SNCA* 遺伝子の重複による家族性パーキンソン病 (PARK4)、*APP* 遺伝子の重複による家族性アルツハイマー病、*SMN1* 遺伝子の欠失による脊髄性筋萎縮症 (SMA) などの疾患が代表的疾患として挙げられるが、これらの変異は通常のエキソーム解析では検出できない。

現在までに、エキソーム解析データを用いて構造変異を検出する解析プログラムは複数報告されている。当科ではこれまで、エキソーム解析で変異が認められないが臨床像から遺伝子変異の存在が強く疑われる症例については、一塩基置換変異や短い挿入欠失変異を見逃している可能性以外に構造変異が存在する可能性を考え、他の手法による解析をその都度追加することで対応してきたが、エキソーム解析を行った全例に対して他の手法を用いて構造変異の有無について確認することは、コスト面などの問題から難しく行えていなかった。本研究では、エキソーム解析後のデータから構造変異を検出する解析パイプラインを構築し、過去当科にてエキソーム解析を行った症例を対象に同パイプラインを用いて解析を行い、神経疾患診断における本解析法の有用性を明らかにすることを目的として検討を行った。

まず、エキソーム解析データに基づいて構造変異を検出するプログラムの一つである Conifer を用いて、エキソーム解析後のデータから構造変異を検出する解析パイプラインを構築した。Conifer へ入力するためのデータ加工、Conifer 出力データへの遺伝子情報などの追加、性染色体アレル数の判定、解析除外症例の選択など、Conifer 以外のデータ処理に関して、複数 (10 以上) のプログラムを作成し、パイプラインを構築した。

続いて、同パイプラインを使用して、これまでに当院当科にて DNA を収集しエキソーム解析を行った症例、および、鹿児島大学大学院医歯学総合研究科神経内科・老年病学講座 高嶋 博 教授、山梨大学医学部神経内科学講座 瀧山 嘉久 教授より DNA を提供され、当科にてエキソーム解析を行った計 1230 例のデータを用いて解析を行った。

結果として、合計 24 症例において原因遺伝子の可能性のある変異が検出され、変異を検出した遺伝子は 11 種類にのぼった (*PMP22*, *SACS*, *SPAST*, *FA2H*, *ZFYVE26*, *OPTN*, *KIF5A*, *LITAF*, *PARK2*, *CYP2U1*, *SBF1* 遺伝子)。これらの構造変異が実際に存在することは、アレイ CGH 法により確認した。24 症例のうち、14 症例については同様の構造変異により疾患を発症した既報告があり、原因遺伝子変異と判断して良いと考えられた。1 症例については、同様の構造変異により疾患を発症した既報告はないものの、遺伝形式が常染色体劣性遺伝で、同一遺伝子に 2 種類の構造変異を有していることから、原因遺伝子変異である可能性が高いと考えられた。また、2 症例については、同様の構造変異により疾患を発症した既報告はないものの、遺伝形式が常染色体優性遺伝形式である疾患であり、原因遺伝子変異である可能性が考えられた。ただし、これら 3 症例について原因遺伝子変異と確定するには、既報告の遺伝子変異を有する症例と臨床的特徴が合致するかどうか、また家系内の発端者以外の遺伝子検査など、追加の確認が必要と考えられた。

7 症例については、遺伝形式が常染色体劣性であるが、検出された構造変異はヘテロ接合性であった。うち 3 症例については健常者集団には認められない rare variant が通常のエキソーム解析で検出されており、原因遺伝子変異の可能性があると考えられた。また、そのうち 1 例については、通常のエキソーム解析結果で既報告のある一塩基挿入変異が検出されており、同変異をホモ接合性に有すると診断されていたが、今回の解析によって欠失変異と一塩基挿入変異の複合ヘテロ接合性変異であると診断が修正された。

また、*SMN1* 遺伝子については偽遺伝子 (*SMN2*) と区別して RPM 値を計算できるよう、Exon 7 および Exon 8 の塩基相違部位に着目した特別な解析アルゴリズムを作成して解析を行い、2 症例において *SMN1* 遺伝子のホモ接合性の欠失を検出した。さらに、疾患との関連性は考えにくいものの、4 例において性染色体トリソミーを検出した (*XXX* 1 例, *XXY* 2 例, *XYY* 1 例)。

本解析パイプラインで検出した構造変異は、これまでのエキソーム解析では見逃されていた変異であり、今回の解析で新たに診断が確定した症例が多くあったほか、過去にエキソーム解析でホモ接合性変異と診断されていたが、今回の解析により欠失変異との複合ヘテロ接合性変異であると診断が修正された症例もあり、本解析法がエキソーム解析による遺伝子診断の正確性を高める上で有用であることが明らかとなった。また、神経内科領域の疾患においても構造変異はまれではなく、一塩基置換変異や短い挿入・欠失変異のみな

らず、構造変異の有無についても評価を行うことが必要であることが今回の研究により明らかとなった。

ただし、今回のエキソーム解析データを利用した解析には以下のような短所が存在する。(1) 含まれるエキソン領域が少ない(1~2個)構造変異は検出困難である。(2) リード数に影響を与える要因が構造変異以外に多数あり、データのバラつきが大きく、精度面での問題が存在する。(3) 多数例の平均値からの解離を利用して判定する方法のため、解析症例中に高頻度で存在する構造変異は検出できない。既存の広範囲を対象として構造変異の検出が可能な手法のうち、最適化された全エキソンを対象とするアレイ CGH 法は、解析対象範囲が同等であり、かつ検出精度の点で劣ると考えられるが、アレイ CGH 法と比較して以下の利点があると考えられる。(1) アレイ CGH 法では得られない、塩基配列情報(一塩基置換変異、短い挿入・欠失変異の有無)が同時に得られる。(2) アレイ CGH 法で解析困難な偽遺伝子の存在する遺伝子についても、条件次第で解析が可能である(*SMNI* 遺伝子など)。

本解析法には、短い構造変異や高頻度で存在する構造変異など原理的に検出できない変異があり、また検出された変異の中には偽陽性も含まれるため、本解析法のみで全エキソンに関して構造変異の有無について確定診断することは難しいと考えられる。従って、コストの問題を考えなければ、エキソーム解析を施行した症例全例について、アレイ CGH 法による解析を併用する方法が最も良いと思われる。ただし、本解析法でも一部の標的遺伝子に対象を限定すれば確定診断的に用いることができる可能性があり、例えば *PMP22* 遺伝子や *SACS* 遺伝子など、構造変異の範囲が一定していて多くのエキソン領域を含むものについては、データのバラつきが非常に少なく、本法のみでも診断可能と思われる。

また、*SMNI* 遺伝子については自作の解析手法を用い、2例において *SMNI* 遺伝子の欠失を検出したが、次世代シーケンサーデータを利用して変異を検出した報告はこれまでにない。なお本解析法は、Exon7, Exon8 の *SMNI* 固有の2塩基のみに着目しており、同部位のホモ接合性の欠失のみ検出可能である。脊髄性筋萎縮症では大多数(~95%)が *SMNI* 遺伝子のホモ接合性の欠失によるとされ、本法により多くの症例で診断が可能と考えられるが、それ以外の変異(複雑な構造変異や、一塩基置換と欠失の複合ヘテロ接合性変異など)は現在の解析法では検出できないため、今後の課題である。

本パイプラインにおけるデータ解析上の問題点は、Coniferによる解析で著しく多い構造変異が検出される症例が少ないながらも含まれていたことである。これらの症例に共通する因子は現時点で明らかにできていないが、サンプルの調整方法からデータ解析までの間に何らかの問題点・相違点が無かったかどうか、検討する必要があると考えられる。

さらに、今回は既報告のある変異についての検討のみしか行えなかったが、既報告の無い遺伝子に特定の疾患において共通する構造変異が無いかなど、全遺伝子を対象とした解析方法についても、今後検討していきたいと考える。

最後に、現在はコスト面から全ゲノム解析を多くの症例に利用することは難しいが、全ゲノム解析に基づく構造変異検出はエキソーム解析と比較して、(1) エキソン領域のみを抽出する作業がないためバイアスが少ない、(2) エキソン領域以外も含むすべての領域のデータが概ね均一に得られる、という非常に大きな利点がある。今後、コストの削減により全ゲノム解析が日常的に行えるようになれば、次世代シーケンサーデータに基づいた構造変異の検出を、より正確に行うことが可能になると考えられる。