


 令和 3 年 3 月 30 日  
 横浜市立大学  
 日本医療研究開発機構

## 新技術ロングリード・シークエンサーで てんかんの原因となる繰り返し塩基配列の正確な配列を決定

横浜市立大学大学院医学研究科 遺伝学 水口剛講師、松本直通教授らの研究グループは、ロングリード・シークエンサー\*1 を用いた新しい解析手法を応用することで、良性成人型家族性ミオクローヌステんかん (Benign adult familial myoclonic epilepsy : 以下、BAFME) の原因となる *SAMD12* 遺伝子の異常伸長配列の正確な配列決定を行うことに成功しました。繰り返し塩基配列の異常伸長(リピート伸長変異)は病気の原因となることが知られており、完全長配列の評価が可能となったことで、正確な分子診断や病態解明に寄与することが期待されます。

この研究は産業医科大学神経内科学講座 豊田知子医師 足立弘明教授、横浜市立大学附属病院遺伝子診療科 宮武聡子講師、東京医科歯科大学難治疾患研究所 三橋里美准教授、横浜市立大学大学院医学研究科神経内科 土井宏准教授 田中章景教授、横浜市立脳卒中・神経脊椎センター 工藤洋祐医師、横浜市立大学附属市民総合医療センター 岸田日帯医師、大和市立病院 林紀子医師、国立病院機構宇多野病院 鈴木理恵医師 木下真幸子医師、信州大学医学部新生児学・療育学 福山哲広医師、昭和大学医学部小児科学講座 加藤光広教授らとの共同研究による成果です。

本研究成果は、英国科学誌『Brain』に掲載されます(日本時間 2021 年 4 月 1 日午前 9 時 1 分掲載)。

### 研究成果のポイント

- 本研究では、Cas9 ターゲットエンリッチメント法\*2 を用いたゲノム上の特定領域(病的なリピート伸長変異)の選択的濃縮と、ナノポア・ロングリード・シークエンサー\*3 を組み合わせることで、リピート伸長変異全長の高精度な配列決定が可能となった。
- 家族性てんかんの原因となる *SAMD12* 遺伝子のリピート伸長変異の配列を完全に決定したことにより、その多様性(繰り返しの回数、配列モチーフ\*4 のパターンとその組成)が明らかになった。
- 非罹患者に存在しない TTTCA リピートの挿入・伸長が良性成人型家族性ミオクローヌステんかん (BAFME) の発症の原因とされていましたが、TTTCA リピートがわずか 14 回の繰り返しである症例が見出され、挿入そのものが重要であることが示唆された。
- 新規技術によるリピートの病原性の理解と診断率の向上につながる画期的な成果である。

### 研究の背景

ヒトゲノムにはおよそ 100 万個の縦列反復配列(タンデムリピート\*5)と呼ばれる塩基の繰り返し配列が存在します。このうち 1~6 塩基の繰り返し配列単位(配列モチーフ)からなるショートタンデムリピートは不安定で変異率が高く、繰り返し回数の異常な増加によりヒト疾患の原因となることが知られています(リピート病\*6)。リピートの伸長変異は、伸長の程度やリピート内配列・構造(配列モチーフのパターンやその組成)が病原性、表現型\*7、リピートの安定性に関連しているという、1 塩基変異には見られない特有の性質があります。例えば、ある種のリピート病ではリピートの繰り返し回数が閾値(いきち)\*8を超えた場合に発症し、リピート数と疾患の発症年齢・重症度との相関が認められています。また、リピート領域内に配列モチーフとは異なる分断配列(Interruption)が存在することがあり、リピート不安定性に影響を及ぼすことも知られています。これらのことから、病的リピート伸長配列を正確に決定することは、リピート病の診断、臨床的マネジメント、病態解明の点で意義が高

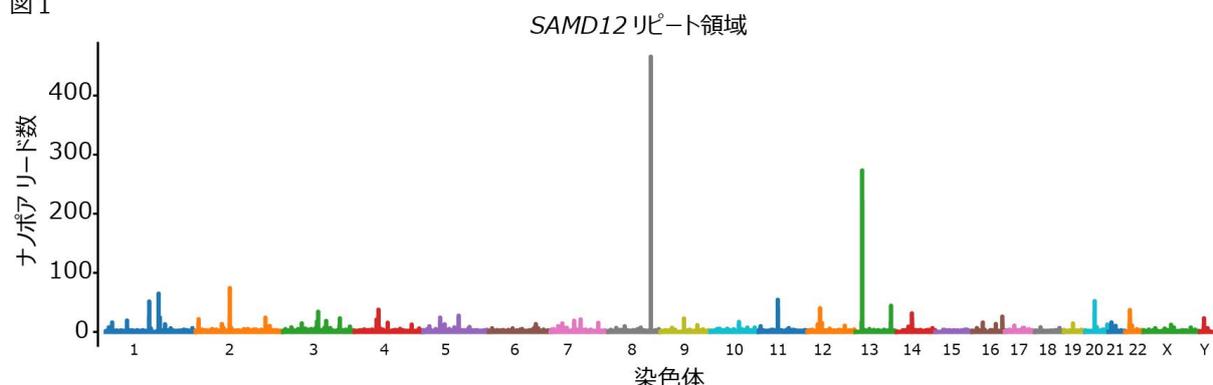
いにも関わらず、その全長を塩基レベルで決定することは従来法では困難でした。従来法では、解析したい領域を PCR 増幅<sup>9</sup>する必要があり、また、解析可能な配列長も 800~1,000 塩基程度ですが、病的なリピート伸長領域の多くは PCR 増幅が困難で、その長さも 1,000 塩基~10,000 塩基を超えることが理由として挙げられます。本研究はこの問題を解決するための取り組みの一つです。

## 研究の内容

良性成人型家族性ミオクローヌステんかん (BAFME) は家族性に発症がみられ、手の細かい震え (振戦様ミオクローヌス) およびてんかん (強直間代発作) を呈する疾患です。SAMD12 遺伝子に存在する TTTTA および TTTCA 配列モチーフの繰り返しの異常な伸長が原因であることが知られています。SAMD12 遺伝子のリピート伸長変異は 2,200~18,400 塩基 (440~3,680 回の TTTTA および TTTCA 配列モチーフの繰り返し) であることが報告されていますが、リピート内の配列・構造についての情報は限定的でした。本研究グループは、平均リード長 (シーケンシングで得られる 1 本 1 本の塩基配列の長さ) が 10,000 塩基を超えるナノポア・ロングリード・シーケンサーを用いることで 25 名の BAFME 患者 (22 家系) のリピート伸長変異配列の解析を行いました。

ロングリード・シーケンサーによって長い配列を解析できるようになると、SAMD12 リピート伸長変異の全長を 1 本のリードで解析することが可能です。しかしロングリード・シーケンサーはエラー率が高いため、精度の高い解析結果を得るためには、リピート配列を深い読み取り深度<sup>10</sup>で繰り返し読むことで信頼性を上げる必要があります。そこで本研究では、シーケンスを行う前に、Cas9 ターゲットエンリッチメント法を用いて SAMD12 遺伝子のリピート領域を含む DNA 断片を濃縮することで、高効率で精度の高い解析を可能にしました。通常、ロングリード全ゲノム解析では全ゲノムに対して 10x~30x の読み取り深度に達するシーケンスデータを解析するのに対し、本研究では SAMD12 遺伝子リピート領域を平均 416x の読み取り深度で解析することができました (図 1)。

図 1



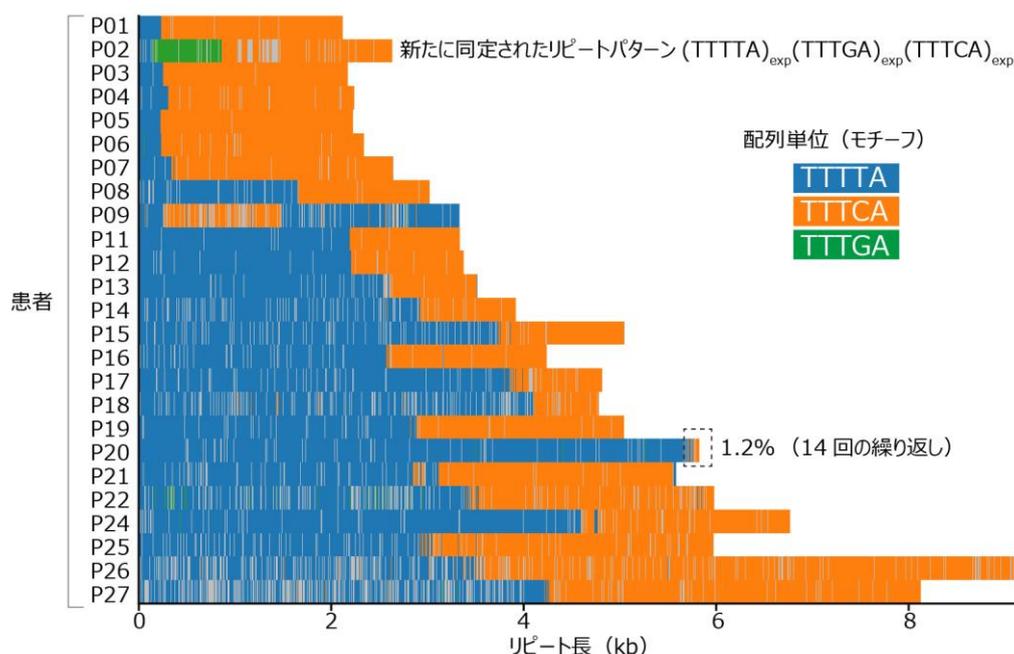
[図 1] ナノポア Cas9 ターゲットエンリッチメント法による SAMD12 遺伝子リピート領域の選択的濃縮の 1 例

8 番染色体 8q24.12 領域に位置する SAMD12 遺伝子のリピート領域が 470 本のナノポアリードで繰り返し読まれており、多くのリードを解析することで配列の信頼性を上げることができる。

コンセンサス配列<sup>11</sup>から明らかになった SAMD12 遺伝子のリピート伸長配列は非常に多様性に富んでおり、大別すると (TTTTA)<sub>exp</sub>(TTTCA)<sub>exp</sub>, (TTTTA)<sub>exp</sub>(TTTCA)<sub>exp</sub>(TTTTA)<sub>exp</sub>, (TTTTA)<sub>exp</sub>(TTTGA)<sub>exp</sub>(TTTCA)<sub>exp</sub> という 3 つのパターンが観察されました (図 2)。さらに、リピート全長に占める TTTCA の繰り返し数の割合には大きな個人差があり (1%~90%)、わずか 14 回の TTTCA 繰り返しで発症している症例も存在しました。わずか 14 回の TTTCA 繰り返しが発症に寄与していたことから、リピート“伸長”というよりも TTTCA リピートの“挿入”そのものが病原性に寄与している可能性が示唆されました。

また、従来法ではリピート伸長変異の存在を明らかにすることが困難であった 2 家系について、病的な伸長配列の全長を決定することで確定診断を行うことができました。

図 2



[図 2] 25 名の BAFME 患者における SAMD12 遺伝子リピート異常伸長配列の多様性

今回解析した BAFME 患者は 2,000 塩基 から 9,000 塩基程度の SAMD12 遺伝子のリピート異常伸長を有している。繰り返し単位 TTTTA, TTTCA, TTTGA をそれぞれ青、オレンジ、緑で表示させている。患者 P2 では  $(TTTTA)_{exp}(TTTGA)_{exp}(TTTCA)_{exp}$  というこれまでに報告のない新たな繰り返しパターンが観察された。また疾患に特異的な変異である TTTCA リピートが全長に占める割合は患者ごとに大きく異なり、患者 P20 では全体の 1.2%、わずか 14 回の繰り返しで発症していることがわかる。

#### 今後の展開

本研究により、従来の手法で解決困難であったリピート伸長変異の完全長配列決定が可能となりました。これにより、これまで複数の断片的な解析結果から総合的に行ってきた変異評価を、本研究手法 1 つで代替、より正確な分子診断が可能となります。また、伸長変異の多様性が明らかになり、今後、臨床症状とリピート内構造（繰り返しの回数、配列モチーフのパターンとその組成、分断配列 Interruption の有無）について塩基配列レベルの比較検討を行うことで、医療情報の提供（発症閾値、重症度、臨床的特徴などとの関連）につながる知見の発見に貢献すると期待されます。

#### 用語説明

\*1 ロングリード・シーケンサー: 10,000 塩基以上の DNA 配列を一つづきに読むことができる装置。オクスフォードナノポア社やパックバイオ社の装置が使われる。

\*2 Cas9 ターゲットエンリッチメント法: CRISPR/Cas9 システムを利用してゲノム中の特定部位に DNA 二本鎖切断を引き起こす。この二本鎖切断部位特異的にシーケンスに必要な配列(アダプター)を付加することで標的とするゲノム特定領域を選択的にシーケンスすることができる。

\*3 ナノポア・ロングリード・シーケンサー: オクスフォードナノポア社が提供するロングリード・シーケンサー。ナノポアというタンパク質の穴が人工膜に埋め込まれており、ナノポアを DNA 分子が通り抜ける時に起きる電流の変化によって、DNA の配列を解析することができる。

\*4 配列モチーフ: リピートを構成する繰り返し配列の単位。

\*5 縦列反復配列 (タンデムリピート): 同じ塩基配列 (TTTCA など) が同方向に連なって繰り返す DNA の繰り返し領域。1~6 塩基を 1 単位とした繰り返し配列 (モチーフ) をショートタンデムリピートと呼ぶ。その繰り返し数に個人差がある。

\*6 リピート病: 特定の塩基配列の繰り返し (リピート) が異常伸長することによって引き起こされる疾患群。現在 40 個以上の疾患が報告されている。

\*7 表現型: 遺伝子の働きによって個体、組織、細胞に現れる形態や機能の特徴。

\*8 閾値: 発症するか、しないか、の境目となる値。一定のリピート数（閾値）を超えた個体のみが発症すると考えられている。

\*9 PCR 増幅 (Polymerase Chain Reaction): 特定の DNA 断片を増幅することができる手法。二本鎖 DNA を高温で 2 本に解離、温度を下げてプライマーを結合させ、DNA ポリメラーゼで相補鎖を伸長させる。このサイクルを 25~40 回繰り返す反応 (PCR) で、特定の DNA 断片を増幅することができる。リピート領域においては、非特異的な配列の繰り返しや高い GC 含量、二次構造によって PCR 増幅が困難な場合がある。また PCR 増幅時に生じるエラーにより正確な配列の評価が困難となる。

\*10 読み取り深度: シークエンサーによって読まれた回数。ゲノム上のある塩基について 20 回重ねて読まれていれば、20x カバレッジであると表わされる。回数を多くできるほど精度の高い解析が可能となる。

\*11 コンセンサス配列: 繰り返し読まれた配列データを用いて、シーケンスの際に生じるエラーを補正することで得られる最も確からしい配列。

## 掲載論文

### **Complete sequencing of expanded *SAMD12* repeats by long-read sequencing and Cas9-mediated enrichment.**

Takeshi Mizuguchi, Tomoko Toyota, Satoko Miyatake, Satomi Mitsuhashi, Hiroshi Doi, Yosuke Kudo, Hitaru Kishida, Noriko Hayashi, Rie S. Tsuburaya, Masako Kinoshita, Tetsuhiro Fukuyama, Hiromi Fukuda, Eriko Koshimizu, Naomi Tsuchida, Yuri Uchiyama, Atsushi Fujita, Atsushi Takata, Noriko Miyake, Mitsuhiro Kato, Fumiaki Tanaka, Hiroaki Adachi, and Naomichi Matsumoto.

*Brain* (2021) DOI : 10.1093/brain/10.1093/brain/awab021

※本研究は、国立研究開発法人日本医療研究開発機構 (AMED) の難治性疾患実用化研究事業「新技術を用いた難治性疾患の高精度診断法の開発」(研究代表者: 松本直通)、厚生労働省、日本学術振興会、武田科学振興財団の支援により実施されました。