

2024年6月18日

国立大学法人東北大学

## 完全長日本人参照ゲノム配列および ミトコンドリアヘテロプラスミーパネルの公開

### 【発表のポイント】

- 日本人のゲノム配列のほぼ全ての塩基配列を決定した日本人参照ゲノム配列 JG3 を公開しました。
- 日本人 8,380 人の全ゲノム解析データを活用し、ミトコンドリア DNA の細胞内多様性を記述したミトコンドリアヘテロプラスミーパネルを公開しました。
- JG3 およびミトコンドリアヘテロプラスミーパネルにより、これまで解析できなかった領域や変異の解析が進み、疾患の機序や原因の解明が期待されます。

### 【概要】

ヒトのゲノム配列、約 30 億塩基対のうちおよそ 27 億塩基対においては、参照ゲノム配列と呼ばれる「ひながた」による解析手法が確立されていますが、残りの 3 億塩基対については解読が困難な領域にありこれまでの日本人参照ゲノム配列でも未解決のままでした。また、一般集団におけるミトコンドリア DNA の多様性はこれまで不明でした。

東北大学東北メディカル・メガバンク機構 (ToMMo) の高山 順准教授らによる研究チームは、今回この残された領域のほとんどを解読し、日本人参照ゲノム配列 JG3 として公開しました。JG3 によりほぼ全てのゲノム配列が解析可能となります。8,380 人のミトコンドリア DNA を解析し、個体内変異の割合をミトコンドリアヘテロプラスミーパネルとして公開しました。ミトコンドリア DNA のバリエーションに起因する疾患メカニズムの解明につながることを期待されます。

なお、今回構築した日本人参照ゲノム配列 JG3 とミトコンドリアヘテロプラスミーパネルは、ToMMo が運用するデータベースである日本人多層オミックス参照パネル (jMorp)<sup>(注1)</sup> で公開されています。

## 【詳細な説明】

### 研究の背景

#### <日本人参照ゲノム配列>

ヒトのゲノム配列は約 30 億塩基対から構成されています。DNA の配列決定法が発明され、ヒトゲノム計画では、2004 年にユークロマチン<sup>(注 2)</sup>領域のおよそ 27 億塩基対が決定されました。しかし、染色体の末端にあるテロメア<sup>(注 3)</sup>領域、また染色体の中心にあるセントロメア<sup>(注 4)</sup>領域、さらにはその周辺に位置するヘテロクロマチン<sup>(注 5)</sup>領域の配列は不明でした。ToMMo でもこれまで日本人の参照ゲノム配列として JG1, JG2, JG2.1 を構築・公開してきましたが(公開当時は基準ゲノム配列と呼称)、これらの領域およそ 3 億塩基対はやはり不明なままでした。

2022 年にヒトゲノム配列の完全決定を目指す国際テロメア to テロメアコンソーシアムから、ヒト完全長ゲノム配列が報告されました。このゲノム配列はヨーロッパ系集団のもので、完全胞状奇胎<sup>(注 6)</sup>と呼ばれる特殊な細胞を利用して作られました。また、様々な DNA 配列決定技術と目視による修正により作成されました。この報告により、ついにヒトゲノム配列の全長を決定できることが技術的に示されました。

そこで ToMMo の研究チームも、日本人の完全長ゲノム配列決定に取り組みました。しかし完全胞状奇胎の入手は困難です。また目視による修正も極めて困難な作業です。そのため国際テロメア to テロメアコンソーシアムの方法をそのまま踏襲することはできません。そこで ToMMo の研究チームは通常の血液検体から、可能な限り自動的に完全長ゲノム配列を決定することに挑みました。

#### <ミトコンドリア DNA>

ヒトの遺伝情報の多くは細胞核の中にあるゲノム配列に含まれていますが、実は細胞の中で呼吸機能を担うミトコンドリアの中にも DNA が含まれています。ミトコンドリアは細胞の中に数百コピー存在しますが、すべて同じ配列であるとは限りません。このように同じ個体、同じ細胞内でもミトコンドリア DNA に多様性があることをヘテロプラスミーと呼びます。そして、病気を引き起こすバリエーションをもったミトコンドリアがある一定の割合を超えると、難聴や神経系の疾患などを引き起こすことが知られています。しかしその一方で一般集団においてミトコンドリア DNA にどのくらいの多様性があるか、すなわち一般集団におけるミトコンドリアヘテロプラスミーの様相は知られていませんでした。

### 今回の取り組み

#### <日本人参照ゲノム配列 JG3>

今回、ToMMo の研究チームは、一般的な血液検体からの完全長ゲノム配列決定に取り組みました。試行錯誤の結果、最新の 2 種類の DNA シーケンシング技術、すなわち、Pacific Biosciences 社の High Fidelity 長鎖リード技術および Oxford Nanopore Technologies 社の超長鎖リード技術を組み合わせることで、ほぼ全ての

塩基配列を決定できることがわかりました。この方針に基づいて3名の血液検体からゲノム配列を解析しました。そうして得られたDNA断片のデータから元のゲノム配列のほぼ全ての塩基配列6組を再構築しました。その後これらの6組のゲノム配列を比較して、異なるところは多数派の配列を採用するようにしました。こうすることで、日本人によくみられる多型を採用したほぼ完全長のゲノム配列を1組構築し、日本人参照ゲノム配列 JG3 と名付けました。

構築したゲノム配列の全長は 3,136,348,502 塩基対でした。構築したゲノム配列にはまだ5,100 塩基対ほどの未決定領域が残っていますが、これまでの3億塩基対の未決定領域に比較すると大幅な進歩です。未決定塩基配列以外の領域においても改善が見られました。たとえば国際参照ゲノム配列 GRCh38 では、突然変異により心不全の原因となる KCNE1 遺伝子やホモシスチン尿症の原因となる CBS 遺伝子が誤って重複しており、この領域の解析が困難となることが知られていますが、JG3 ではこれらの誤りが修正されています。

#### <ミトコンドリアヘテロプラスミーパネル>

今回、ToMMo の研究チームは、東北メディカル・メガバンク計画<sup>(注7)</sup>のコホート参加者のうち8,380人のゲノムデータを詳細に調べ、日本人集団におけるミトコンドリアヘテロプラスミーの様相をはじめて明らかにし、パネルとして公開しました。

解析の結果、ミトコンドリアヘテロプラスミーは稀ながらも一般集団で見られることがわかりました。またヨーロッパ系集団とはミトコンドリアの系統(ハプログループ<sup>(注8)</sup>)が異なるものの、ヘテロプラスミーの観察される場所は極めてよく似ていることが明らかになりました。このことは遺伝的な由来によらずヘテロプラスミーが生じる共通のメカニズムがあることを示唆しています。

今回構築した日本人参照ゲノム配列 JG3 とミトコンドリアヘテロプラスミーパネルは、ToMMo が運用するデータベースである日本人多層オミックス参照パネル(jMorp)で公開されています。

【jMorp】

サイト名: Japanese Multi Omics Reference Panel (jMorp)

言語: 英語

URL: <https://jmorp.megabank.tohoku.ac.jp/>



#### 今後の展開

##### <日本人参照ゲノム配列 JG3>

参照ゲノム配列は短鎖型次世代シーケンサー<sup>(注8)</sup>を用いたゲノム解析における「ひ

ながた」の役割を果たします。JG3 はこれまで解析が不可能だった領域の解析を可能にし、またより精度の高いゲノム解析に貢献することが期待されます。

<ミトコンドリアヘテロプラスミーパネル>

ミトコンドリア病に代表されるミトコンドリア DNA のバリエーションに起因する疾患メカニズムの解明につながることを期待されます。

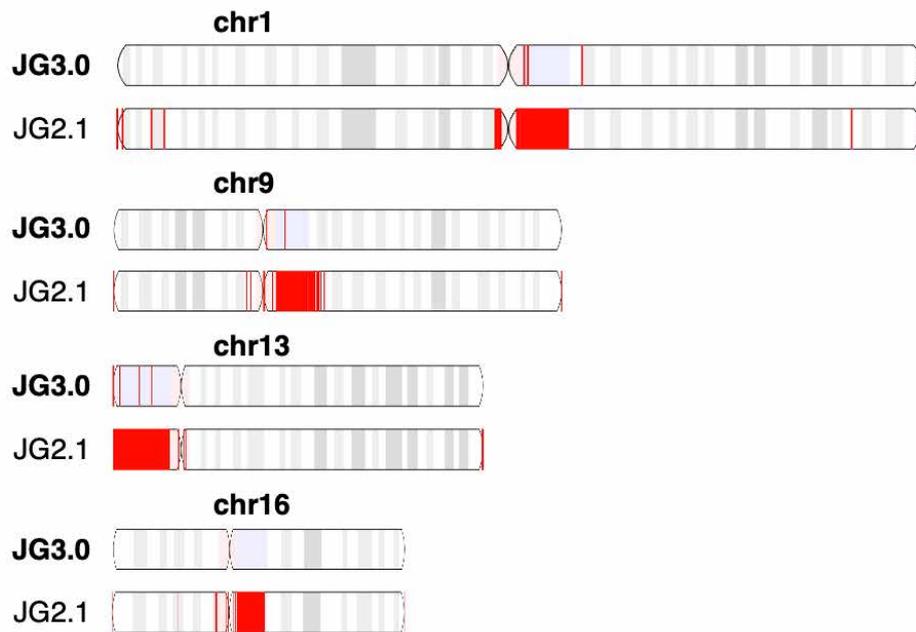


図 1. JG3 と JG2.1 の比較。赤色は配列未決定の領域を示しています。JG3 では未決定領域が大幅に縮小されました。

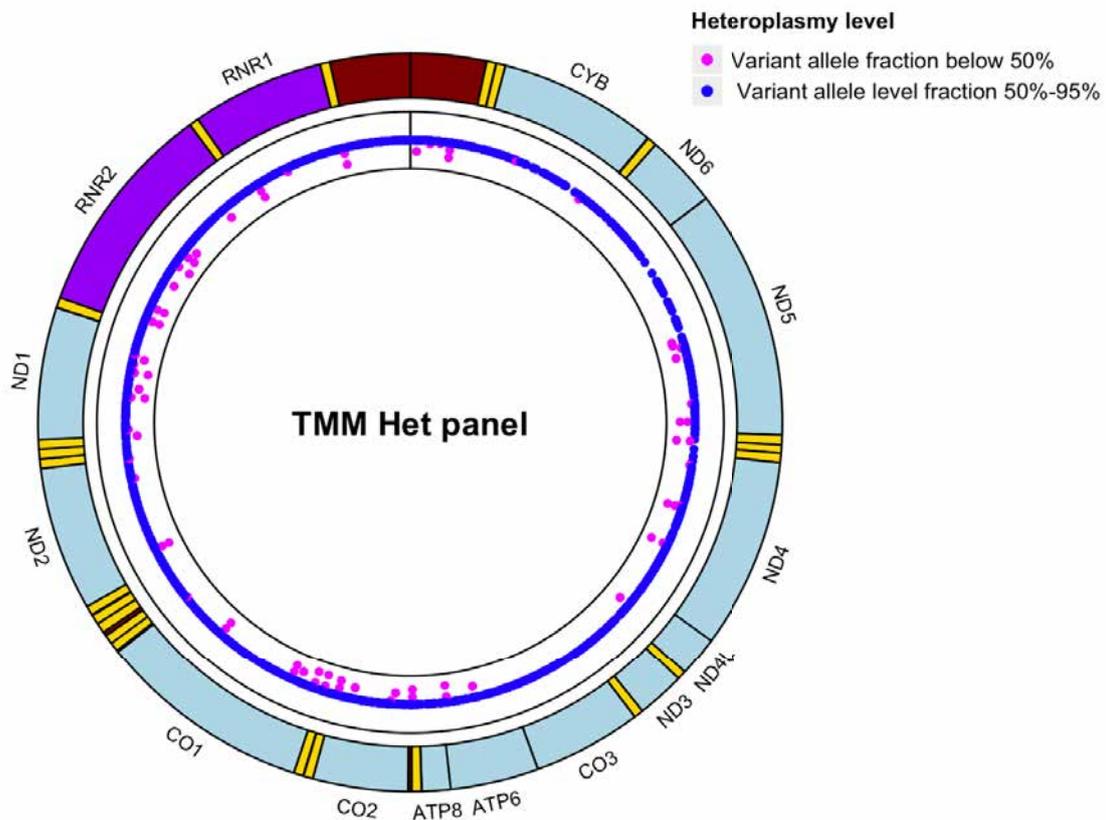


図 2. ミトコンドリアヘテロプラスミーの概要。図は環状のミトコンドリア DNA を円形に表したものです。外側のトラックに遺伝子の領域とその名前を示しています。内側のトラックにレベルに応じて色分けしたヘテロプラスミーサイトを示しています。

**【謝辞】**

本研究は AMED ゲノム医療実現バイオバンク利活用プログラム事業「大規模ゲノム解析に必要な計算基盤構築とゲノム解析に関する研究 (JP21tm0424601)」の助成を受け実施されました。

**【用語説明】**

注1. 日本人多層オミックス参照パネル (jMorp: Japanese Multi Omic Reference Panel): TMM 計画のコホート調査によって得られた試料を解析した結果を、個人識別性のない頻度情報等に行っている公開データベース。  
<https://jmorp.megabank.tohoku.ac.jp/>

注2. ユークロマチン:ゲノム DNA とヒストンと呼ばれるタンパク質の複合体をクロマチンと呼び、細胞周期や転写などの細胞の状態に応じて凝集したり弛緩したりすることが知られている。ユークロマチンはクロマチンがもっぱら弛緩している領域で、比較的転写が活発な領域であり、DNA 配列決定の観点からは、比較的解読しやすい領域である。

- 注3. テロメア:染色体の末端に位置する短い単位のリピート配列からなる領域。細胞分裂時の DNA の複製が不完全なために短くなることから老化のマーカーとして知られる。DNA 配列決定の観点からは、解読困難な領域である。
- 注4. セントロメア:染色体の中央に位置する短い単位のリピート配列からなる領域。細胞分裂時には染色体を 2 つの新しい細胞に均等に分けるために重要な領域である。DNA 配列決定の観点からは、解読困難な領域である。
- 注5. ヘテロクロマチン:クロマチンが(ほとんど常に)凝縮しており転写が不活発な領域。1 番染色体や 9 番染色体にはセントロメア周囲に巨大なヘテロクロマチン領域が存在する。DNA 配列決定の観点からは、解読困難な領域。
- 注6. 完全胞状奇胎:受精時の異常により精子由来のゲノム DNA が倍化して作られた細胞。通常の細胞は父母両親由来のゲノムを有するが、完全胞状奇胎は父由来のゲノムのみを持つ。そのためゲノム配列決定が行いやすい。
- 注7. 東北メディカル・メガバンク計画:日本最大規模の一般住民ゲノム・コホート調査を実施しており、次世代医療の実現に貢献するため、個人のゲノム情報に紐づく多様なデータから複合バイオバンクを構築し、長期追跡している。
- 注8. ハプログループ:ミトコンドリア DNA の多様性をもとにグループ分けしたもの。ヨーロッパ系やアフリカ系集団には珍しく日本人には多いハプログループにはハプログループ G や M が知られている。
- 注9. 短鎖型次世代シーケンサー:全ゲノム解析の主流となる解析技術で、150塩基対程度のリードと呼ばれる短い断片の単位で並列に数億回読んだデータを産生できる。検体のゲノム多型を推定するためにはリードを参照ゲノム配列に当てはめる。

**【問い合わせ先】**

(研究に関すること)

東北大学 東北メディカル・メガバンク機構

ゲノム遺伝統計学分野 准教授

高山 順(たかやまじゅん)

TEL: 022-274-2371

Email: [jtakayama@tohoku.ac.jp](mailto:jtakayama@tohoku.ac.jp)

(報道に関すること)

東北大学 東北メディカル・メガバンク機構

広報戦略室長

長神 風二(ながみ ふうじ)

TEL: 022-717-7908

Email: [tommo-pr@grp.tohoku.ac.jp](mailto:tommo-pr@grp.tohoku.ac.jp)