

東北大学東北メディカル・メガバンク機構

長鎖リードシーケンサーによる  
ゲノム構造多型の解析・公開  
長い情報を読んで初めてわかるゲノムの正体

【発表のポイント】

- 日本で初めて一般住民の集団を対象としたゲノム構造多型<sup>\*1</sup>の同定を実施。同定された多型が親子間で継承されていることを指標に、ゲノム構造多型を高い精度で検出できていることを確認した。
- 東北メディカル・メガバンク機構 (ToMMo) が樹立・保管している細胞試料と最新型長鎖リードシーケンサー<sup>\*2</sup>を活用することで、精度の高いゲノム構造多型の解析が実現した。
- 解析した構造多型の頻度情報を公開することで広く利活用可能とした。ゲノム構造多型が関連する疾患研究において、日本人の貴重な比較参照データとして役に立つことが期待される。

【概要】

ヒトゲノムには小さなものから大きなものまで、様々なバリエーション<sup>\*3</sup>が存在します。これまで ToMMo では、約 5 万人のゲノム解析を通して、比較的小さなサイズのバリエーションの位置や頻度を収載したデータベースを構築・公開してきました。今回の研究では、最新型の長鎖リードシーケンサーを用いて 333 人の全ゲノム解析を行うことで、50 塩基対を超える大きな挿入・欠失 (構造多型) を詳細に解析しました。特に、ToMMo が樹立・保管している細胞試料を活用することで、安定した品質の長鎖リードシーケンシング解析を可能にしました。また、今回の 333 人はトリオ (成人とその両親からなる 3 人組) の組み合わせで構成されており、親子間の解析結果をもとに「答え合わせ」をすることで精度検証も行い、高い精度でゲノム構造多型を検出できていることを確認しました。本論文の成果は日本人構造多型参照パネル JSV1 として、公開データベース jMorP で公開されています。今回の解析結果は、小さなサイズのバリエーション情報だけでは分からなかった疾患の原因探索やがんゲノム解析への応用に有効であると期待されます。

このゲノム構造多型解析結果「JSV1」の構築についての論文が 2022 年 9 月 20 日付で *Communication Biology* 誌に掲載されました。

## 【背景】

ToMMo ではこれまで「短鎖リードシーケンサー<sup>\*2</sup>」を用いて、5 万人の一般住民のゲノムを解析し、多数の SNV<sup>\*3</sup> や小さな挿入・欠失 (INDEL)<sup>\*3</sup> の情報をリファレンスパネルとして公開してきました。一方、このような小さなバリエーション以外にも、ヒトゲノムには特に構造多型と呼ばれる大きなバリエーションも多数存在し、個人の体質や疾患発症に関わっていると考えられています。しかし、「短鎖リードシーケンサー」による解析では構造多型を精確に捉えることはできず「長鎖リードシーケンサー」による解析が必要でした。さらに長鎖リード技術による安定した全ゲノム解析を行うためには、いかに大量の高品質ゲノム DNA を準備するかが課題でした。

## 【詳細】

### 高品質ゲノム DNA の取得

ToMMo が樹立・保管している培養細胞<sup>\*4</sup> (増殖 T 細胞) から、抽出直後の断片化の少ない高品質ゲノム DNA を得ることができました。また増殖 T 細胞は試験管内で増やすことができるため、バイオバンク<sup>\*5</sup> に保管されている DNA 試料を消費することなく、大量のゲノム DNA を新たに取得することができました。増殖 T 細胞は、他のゲノムプロジェクトで用いられることの多いリンパ芽球様細胞株 (LCL: Lymphoblastoid Cell Line) よりも短期間で効率良く樹立することのできる細胞で、東北メディカル・メガバンク計画バイオバンクの特色となっています。

### 長鎖リードシーケンサー (Oxford Nanopore PromethION) による解析

高品質ゲノム DNA により、安定した品質の長鎖リードシーケンス解析が可能となり 333 人の解析を実施しました。特に、ゲノム DNA を最適な条件で均一に断片化することで、検体ごとのリード長や得られるデータ量のばらつきを少なくできるよう工夫しました。一般住民集団を対象とした構造多型解析としては日本初の成果です。

### トリオを用いた精度検証

今回の 333 人は、家系情報付きである三世代コホート調査の長所を活かし、すべてトリオ (成人とその両親からなる 3 人組) の組み合わせでシーケンス解析を行っています。子にあたる成人が、両親の遺伝子を受け継いでいることを利用して精度検証を行い、高い精度でゲノム構造多型が検出できていることを報告しました。

### データベースの構築と公開

この研究で同定した日本人における様々な構造多型の位置・頻度は、2021 年 12 月に jMorp 上で公開されています。公開情報では、一般集団のゲノム多型の頻度情報を計算するため偏りがないう血縁関係のある検体を除き、両親 (合計 222 人) のみを対象としています。見つかった構造多型それぞれに、上記の精度検証結果を付与し

ています。一般住民集団を対象に長鎖リード解析を行い、構造多型情報を公開した事例は日本初、海外でも数例認められるのみです。

### <jMorp>

公開データベース日本人多層オミックス参照パネル。東北メディカル・メガバンク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して公開している。

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

言語: 英語

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



### 【今後の展望】

今回の長鎖リードシーケンサーによるゲノム構造多型の解析・公開により、SNV や短い INDEL の情報だけでは分からなかった、日本人における疾患の原因探索やがんゲノム解析への応用が期待できます。今後は、解析対象人数を増やすことで頻度の低い構造多型まで網羅するとともに、重複や逆位といった様々なパターンの構造多型解析への拡張を検討しています。

【謝辞】日本医療研究開発機構 (AMED) 東北メディカル・メガバンク計画 (東日本大震災復興特別会計) 事業「東北メディカル・メガバンク計画 (東北大学) 東日本大震災復興特別会計分 (JP17km0105001)」、AMED ゲノム医療実現バイオバンク利活用プログラム (東北メディカル・メガバンク計画) 事業「東北メディカル・メガバンク計画 (東北大学) (JP21tm0124005)」、文部科学省先端研究基盤共用促進事業などの支援を受けて実施されました。

### 【用語解説】

- \*1 ゲノム構造多型 : ゲノム配列において、SNV (後述) や短鎖リードシーケンサーで検出できるような INDEL (後述) などの短い長さのバリエーションではなく、数十から数千、あるいはそれ以上の塩基が個人間で異なる多様性のこと。
- \*2 長鎖 (短鎖) リードシーケンサー : いずれも大量のゲノム情報を同時並列で高速に解析可能な装置。ゲノム DNA 断片を数百塩基単位で解析するのが短鎖リードシーケンサー、数千から万単位の塩基を解析可能なのが長鎖リードシーケンサーである。短鎖リードシーケンサーは、リードあたりの解析精度やコスト面の優位性を活かして小さなバリエーションの解析に適している。長鎖リードシーケンサーは、

その長いリードで大きなバリエントを覆うことができるため、構造多型の解析に適している。

- \*3 バリエント/SNV/INDEL : それぞれ個人間で異なるゲノム領域。このうち、一塩基のみの置換を一塩基バリエント(SNV)、比較的少数の塩基の挿入や欠失をINDEL、50 塩基対以上の様々なバリエントを構造多型(SV)とよぶ。本論文では、構造多型のうち、挿入・欠失を解析対象とした。
- \*4 ToMMo が樹立・保管している培養細胞 : ToMMo では得られた血液試料から、2種類の培養可能な細胞株を樹立・保管している。ひとつはリンパ芽球様細胞株(LCL) とよばれるもので、他のゲノムプロジェクトでも用いられている。もうひとつは増殖 T 細胞で、LCL よりも短期間で効率良く樹立することができる。
- \*5 バイオバンク : 生体試料を収集・保管し、研究利用のために提供を行う取り組み。東北メディカル・メガバンク計画のバイオバンクは、コホート調査の参加者から血液・尿などの生体試料を集めている。

#### 【論文情報】

論文名 : Construction of a trio-based structural variation panel utilizing activated T lymphocytes and long-read sequencing technology

(増殖 T 細胞と長鎖リードシーケンシング技術を用いて、トリオ解析に基づく構造多型参照パネルを構築した)

著者名 : Akihito Otsuki<sup>#</sup>, Yasunobu Okamura<sup>#</sup>, Noriko Ishida, Shu Tadaka, Jun Takayama, Kazuki Kumada, Junko Kawashima, Keiko Taguchi, Naoko Minegishi, Shinichi Kuriyama, Gen Tamiya, Kengo Kinoshita, Fumiki Katsuoka, Masayuki Yamamoto

# Equally contributed

大槻晃史、岡村容伸、石田典子、田高周、高山順、熊田和貴、川嶋順子、田口恵子、峯岸直子、栗山進一、田宮元、木下賢吾、勝岡史城、山本雅之

雑誌名 : Communications Biology

DOI: 10.1038/s42003-022-03953-1

公表日 : 2022 年 9 月 20 日

**【お問い合わせ先】**

(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

基盤情報事業部長

木下 賢吾(きのした けんご)

電話番号:022-274-5952

Eメール:[jmorp@omics.megabank.tohoku.ac.jp](mailto:jmorp@omics.megabank.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

長神 風二(ながみ ふうじ)

電話番号:022-717-7908

ファクス:022-717-7923

Eメール:[pr@megabank.tohoku.ac.jp](mailto:pr@megabank.tohoku.ac.jp)