

2022年6月30日

東北大学東北メディカル・メガバンク機構  
東北大学高等研究機構未来型医療創成センター  
日本医療研究開発機構

## 日本人5万人分の全ゲノム情報の解析を完了 ～3.8万人分の全ゲノムリファレンスパネルを公開～

### 【発表のポイント】

- ・ 日本人一般住民集団5万人分の全ゲノム情報<sup>\*1</sup>の解析を完了しました。
- ・ 官民共同10万人全ゲノム解析計画に基づくものであり、2021年3月に設立した「全ゲノム情報と医療・健康情報の統合解析コンソーシアム<sup>\*2</sup>」からも大きな支援を受けています。解析したデータは、順次、全国の研究者に提供します。
- ・ 5万人のうち3.8万人分の解析データをもとに、日本人全ゲノムリファレンスパネル<sup>\*3</sup> 38KJPNを公開しました。

### 【概要】

東北大学東北メディカル・メガバンク機構(ToMMo)は、官民共同10万人全ゲノム解析計画に基づき、東北メディカル・メガバンク計画(TMM計画:参考1)によるコホート調査<sup>\*4</sup>に参加した約15万人のうち5万人分の全ゲノム解析を完了しました。一般住民の集団を対象とした大規模な全ゲノム解析としては世界有数の規模であり、今後、日本における革新的な創薬をはじめ多様な研究開発に貢献することが期待されます。今回の解析をもとに、血縁関係にないと推定される3.8万人分のバリエーション<sup>\*5</sup>のアレル・ジェノタイプ頻度<sup>\*6</sup>情報からなる日本人全ゲノムリファレンスパネル 38KJPNがデータベース jMorp(参考2)で公開されます。

今後も順次、データを公開し全国の研究者に提供します。

## 【詳細】

### <背景>

一般住民を対象に全ゲノム解析を行うことは、革新的な創薬をはじめ多様な研究開発にとって極めて重要であり、日本製薬工業会(製薬協)の「製薬協政策提言 2021<sup>\*7</sup>」でも大規模な実施の必要性が言及されています。

ToMMo は 2013 年 5 月に開始したコホート調査で提供された生体試料を解析した大規模なデータベースを構築し、同年 11 月に 1,000 人分の全ゲノム解析の完了<sup>\*8</sup>を発表したことを皮切りに、着々と解析数を積み上げ、2021 年 12 月に約 14,000 人分の解析結果をもとにした日本人全ゲノムリファレンスパネル 14KJPN<sup>\*9</sup>を発表しました。

一方、先進各国においてはさらに大規模な取り組みがなされており、最近では、2021 年 11 月に英国の UK Biobank が 20 万人分の全ゲノム解析情報を、翌年 3 月には米国の All of US が 10 万人分をそれぞれ公開したこと<sup>\*10</sup> が大きな話題となりました。

こういった背景から ToMMo は全ゲノム情報と医療・健康情報の統合解析コンソーシアムの支援も受け、大規模な全ゲノム解析を推進していました。

### <内容>

#### ◆5 万人分の全ゲノム解析

ToMMo では実験チーム・情報解析チームが一丸となり全ゲノム解析を進めてきたことで、2022 年 1 月末日に約 5 万検体の全ゲノム情報の基礎的な解析を完了しました。この 5 万検体分のデータを集約・統合解析を行い、5 万人分の全ゲノム情報が含まれるデータベースを作成しました。これに含まれる各検体の品質情報等は jMorp ウェブサイトの Repository (<https://jmorp.megabank.tohoku.ac.jp/repository>) からご覧いただけます。

#### ◆日本人全ゲノムリファレンスパネル 38KJPN

5 万人分の全ゲノム情報から、頻度が偏らないよう血縁関係にないと推定される 3.8 万人を抽出し、日本人全ゲノムリファレンスパネルを構築しました。一つ前のバージョンとなる 14KJPN 全ゲノム参照パネルと比較すると検体数は約 2.7 倍のスケールとなります。

38KJPN・14KJPN に収録される SNV<sup>\*11</sup>・INDEL<sup>\*12</sup> (常染色体)の数は以下の通りです。38KJPN と 14KJPN の SNV・INDEL 数を比べると、それぞれ約 7,000 万個・860 万個増加し、より多くのバリエントを収載することができました。

全ゲノムパネル	SNV	INDEL
38KJPN	176,329,248	21,824,946
14KJPN (過去バージョン)	106,705,823	13,130,321

今回のリリースでは常染色体上の SNV・INDEL アレル・ジェノタイプ頻度情報を公開します。なお X 染色体、およびミトコンドリアについては後日公開予定です。

38KJPN のアレル・ジェノタイプ頻度情報は jMorp ウェブサイト(参考 2)からダウンロード可能です。(ジェノタイプ頻度情報のダウンロードにあたっては、ORCID<sup>\*13</sup> と連携する認証を行い、データ移転契約 (DTA:Data Transfer Agreement) をご確認いただく必要があります。)

#### ◆日本人 HLA 遺伝子<sup>\*14</sup> パネル 38KJPN-HLA

ヒトの HLA 遺伝子は多数のアレルを持つことが知られています。HLA 遺伝子は SNV や INDEL を対象とした解析方法により正確に遺伝子型推定を行うことは困難です。そこで、38KJPN を構成する検体に対し HLA 遺伝子に特化した解析方法を適用することで、HLA 遺伝子の遺伝子型を集めた HLA 参照パネルを作成しました。

#### 【今後の展望】

ToMMo は 1.4 万人の解析完了から約半年で 5 万人の全ゲノム解析を達成しました。今後も引き続き解析を続け、目標とする 10 万人分のゲノム解析を完遂し、英国、米国のような大規模ゲノム・データ基盤を持つことを目指しています。この基盤を全国の研究者が利活用することにより、多因子疾患に有効な健康リスク予測による個別化予防や、革新的な創薬開発の早期実現が可能になります。これは日本人だけでなく広く東アジアにも適用可能な基盤と予測されます。

10 万人の全ゲノム解析が完了すれば、全ゲノムリファレンスパネルもさらに大規模にすることができ、バリエーションの頻度をより正確に表すことができます。

#### 【用語説明と参考資料】

\*1 全ゲノム情報 : ここでは次世代シーケンサーにより解析した個人のすべてのゲノム情報を指す。

\*2 全ゲノム情報と医療・健康情報の統合解析コンソーシアム : 2021 年 3 月にスタートした ToMMo と製薬企業 5 社の参画によるコンソーシアム。(プレスリリース「10 万人の全ゲノム解読を目指した、統合解析コンソーシアムの設立」<https://www.megabank.tohoku.ac.jp/news/45027>)

\*3 全ゲノムリファレンスパネル : 東北メディカル・メガバンク計画で実施された、日本人の一般住民数千人の全ゲノム次世代シーケンシング解析により、検出されたゲノム DNA バリエーションから構築された日本人ゲノム配列のパネル。一般住民のゲノム情報をもとにしているためリファレンスとして有効である。

\*4 コホート調査 : 一定の集団を対象に長期に渡って生活習慣と疾患の発症について調査することで、生活習慣または特定の事象の暴露と疾患発症リスクとの関係を解析するための分析疫学の一手法。TMM 計画のコホート調査ではゲノム解析を実施しており、ゲノム情報(体質)と生活習慣や環境がどのように病気などの表現型と関

連するか調べることができる。

\*5 バリエント：標準となるゲノム配列とは異なる箇所のこと。

\*6 アレル・ジェノタイプ頻度：アレル頻度は、ある集団における DNA バリエントの塩基 (A,T,G,C) の頻度で、アレル (同じ座位上で対立して存在する塩基) ごとに算出する。今回は対象となった日本人 3.8 万人中の頻度となり、最大で 7.6 万アレルのうちにどれだけ検出されたか計算される。ジェノタイプ頻度は遺伝子型頻度ともよばれ、父母から由来する二つのアレルの組み合わせの頻度。今回の発表では、対象となる 3.8 万人の中で、ホモで持つ (父母由来の情報が双方ともある)、ヘテロで持つ (父母いずれかからのみ持つ) などを分けて算出している。

\*7 製薬協政策提言 2021：

[https://www.jpma.or.jp/vision/policy\\_recommendations2021.html](https://www.jpma.or.jp/vision/policy_recommendations2021.html)

\*8 1,000 人分の全ゲノム解析の完了：プレスリリース「疾病原因探索の基盤となる 1000 人分の全ゲノム配列の高精度解読を完了」

<https://www.megabank.tohoku.ac.jp/news/682>

\*9 14KJPN：プレスリリース「ゲノムリファレンスパネルとゲノム解析情報を拡充」

<https://www.megabank.tohoku.ac.jp/news/47145>

\*10 先進各国の発表：UK Biobank の発表「Whole Genome Sequencing data on 200,000 UK Biobank participants made available for research」

<https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/whole-genome-sequencing-data-on-200-000-uk-biobank-participants-available-now>

All of US の発表「NIH's All of Us Research Program Releases First Genomic Dataset of Nearly 100,000 Whole Genome Sequences」

<https://allofus.nih.gov/news-events/announcements/program-releases-first-genomic-dataset>

\*11 SNV：一塩基バリエント。ゲノム配列において、ある領域で DNA の塩基配列が個人間で一塩基のみ異なる多様性のこと。

\*12 INDEL：ゲノム配列における塩基配列の挿入 (insertion) または欠失 (deletion) のどちらかあるいは両方。

\*13 ORCID：研究者等学術的な著作の著者を一意的に識別するために作られた英数字コード。

\*14 HLA 遺伝子：ヒト白血球型抗原 (Human Leukocyte Antigen) 略して HLA はヒトの免疫に関わる重要な分子であり、HLA 遺伝子は HLA を構成する各抗原をコードする遺伝子。HLA を構成する各抗原は多くの「型」を持っており、この組み合わせは「HLA 型」と呼ばれる。異なる HLA 型を持つ細胞は免疫系により異物として攻撃されるため、造血幹細胞移植や臓器移植ではドナーとレシピエント間における HLA 型の適合度が重要視される。

【参考 1】<東北メディカル・メガバンク計画>

東北メディカル・メガバンク計画は、東日本大震災からの復興事業として平成 23 年度から始められ、被災地の健康復興と、個別化予防・医療の実現を目指しています。

ToMMo と岩手医科大学いわて東北メディカル・メガバンク機構を実施機関として、東日本大震災被災地の医療の創造的復興および被災者の健康増進に役立てるために、合計 15 万人規模の地域住民コホート調査および三世代コホート調査を平成 25 年より実施し、収集した試料・情報をもとにバイオバンクを整備しています。

2015 年度より、日本医療研究開発機構 (AMED) が本計画の研究支援担当機関の役割を果たしています。

### 【参考 2】<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/>



#### 【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

(報道担当)

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)

(AMED 事業に関すること)

日本医療研究開発機構 (AMED)

ゲノム・データ基盤事業部 ゲノム医療基盤研究開発課

電話番号: 03-6870-2228

E メール: [tohoku-mm@amed.go.jp](mailto:tohoku-mm@amed.go.jp)