



令和元年9月2日

報道機関 各位

東北大学東北メディカル・メガバンク機構
国立研究開発法人日本医療研究開発機構

ゲノム・オミックス解析情報の公開データベース jMorp の収載データを大幅拡充 ～メタボローム解析データを1万5千人に拡大し縦断解析データも格納、 全ゲノム解析データが4千7百人に～

【発表のポイント】

- ・ 東北大学東北メディカル・メガバンク機構 (ToMMo) は、疾患の比較対照として有効な公開データベース日本人多層オミックス参照パネル (jMorp) の収載データを大幅に拡充しました。一般住民の解析情報である jMorp の収載データと、疾患由来のデータを比較することなどで、疾患の原因や治療法の発見につながります。
- ・ jMorp のメタボローム解析*1対象者数を1万人から約1万5千人に拡張し、さらに経時的な変化データを初めて追加しました。また、より精密かつ正確に疾患患者サンプル等と比較可能な定量値データを新たに追加しました。
- ・ 全ゲノム解析データを約1千人分増やして約4千7百人とし、全ゲノムリファレンスパネル*2 4.7KJPN としました。

【概要】

jMorp は 2015 年 7 月に ToMMo が世界で初めて 500 人以上の血漿に対する網羅的メタボローム及びプロテオーム*3 の統合解析結果を公開したデータベースです。最初の公開以来着々と収載データを増やすとともにゲノム解析の情報を追加し、現在ではゲノミクス、プロテオミクス、メタボロミクスというヒトに関わる生命科学の総合的な情報を網羅的に収載するリファレンスパネル*4となっています。jMorp の情報は主に東北メディカル・メガバンク計画 (【参考】を参照) の長期健康調査に参加された方々のご協力のもとに収集されています。

今回、この jMorp のメタボローム解析対象者数を 10,000 人から約 15,000 人に拡張しました。さらに詳細二次調査*5 で得られた、1,000 人分のメタボローム解析結果を加え、3～5 年の加齢によりメタボロームがどのように変化するかも比較可能としました。そして新たに定量値の解析結果を追加しました。今回の拡充により、さらに精密かつ正確なメタボローム解析が可能なリファレンスパネルとなりました。

全ゲノム解読情報に基づくアレル頻度*6 情報についても、データのもととなる解析人数を 3,500 人から約 4,700 人 (4.7KJPN) としました。今回のアレル頻度パネル拡張では、東北

メディカル・メガバンク計画による宮城県と岩手県でのコホート調査*7への協力者の他に、外部コホートへの協力者の検体を追加し、より日本人として代表性の高いゲノム情報となりました。

今回の拡充により、4.7KJPN は日本人のゲノムデータを有するデータベースとして最大であり、約 15,000 人分のメタボローム情報は世界で最大級の収載数となりました。また、メタボロームの経時変化データの公開は世界初です。

全国の研究者が jMorp のデータを用いることにより、疾患の原因や予防法、バイオマーカー*8の発見などにつながる研究の進展が期待されます。

本研究は、国立研究開発法人日本医療研究開発機構（AMED）による東北メディカル・メガバンク計画のもと東北大学東北メディカル・メガバンク機構によって行われています。

【詳細】

■jMorp の経緯と機能

jMorp は、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たデータベースです。2015年7月の公開以来、2019年7月末までに95カ国から延べ20,715のサイト訪問、148,620のページ閲覧実績があり、広く研究者に利用されています。

当初、500人分の網羅的メタボロームとプロテオームの統合解析を行った世界初の成果として公開した jMorp ですが、これに別サイトにて公開していた、全ゲノム解読情報に基づくアレル頻度情報、日本人基準ゲノム配列*9を加え、それぞれ徐々に収載人数や公開範囲を増やし、現在日本人の生命科学データの総合的なリファレンスパネルとして他に類のない基盤となっています。

表 1：jMorp の経緯

時期	アップデート内容
2014年8月	1KJPN（1,000人分の全ゲノムリファレンスパネル）公開。（別サイト）
2015年7月	jMorp 初公開 。500人分の網羅的メタボローム及びプロテオーム統合解析結果を公開
2016年6月	2KJPN 公開（別サイト）
2016年8月	メタボロームの解析人数を 1,000 人に拡大。項目間関連情報・ペプチド情報を追加
2017年7月	3.5KJPN 公開（別サイト）
2017年10月	メタボロームの解析人数を 5,000 人に拡大。年齢別の代謝物濃度分布・ネットワーク解析結果を追加
2018年6月	3.5KJPNv2 を jMorp に統合して公開。X 染色体とミトコンドリアゲノム情報を追加
2018年8月	メタボロームの解析人数を 10,000 人に拡大

■メタボローム解析の拡充内容

解析人数、対象とする代謝物の種類を拡充、さらに同じ対象者集団についての経時的な情報を追加し、大規模・高精度・多彩な代謝物の濃度分布情報を含むデータベースとなりました。

1) 【データ拡充】NMRによるメタボローム解析の代謝物の種類と解析人数を拡充：
NMRメタボローム解析^{*10}では2018年に37代謝物について、10,719人分の定量値・分布情報を公開しています。この定量解析系に今回、グリシン合成、クエン酸経路などに属する6種類の代謝物を新規に追加しました。この43代謝物についての定量解析法を2018年度jMorp公開データに適用し再解析した結果に加え、さらに今回新たに5,000人分の解析結果を追加し、約15,000人という大規模なものへと拡張しました。また、中には初めて公開する妊娠中あるいは産後の代謝物のデータ約1,600人分も含まれます。

2) 【新規】MSによる標的メタボローム解析に定量情報を追加：

MSメタボローム解析^{*11}では、高速液体クロマトグラフ三連四重極型質量分析装置(LC-MS/MS)による標的メタボローム解析に新たに代謝物キット解析^{*12}を導入し、血漿10 μ Lから検出された110代謝物を新規に追加しました。なお解析対象者数は、約1,500人分です。この110代謝物の新規データは、これまで公開してきた相対値情報ではなく、他の研究機関での解析結果と比較可能な定量値情報として提供します。MSメタボローム解析については、これまで施設間での解析データを直接比較することは難しいと言われていましたが、今回の発表はこれを解決するものです。また、ガスクロマトグラフ三連四重極型質量分析装置(GC-MS/MS)による代謝物についても、新たに約1,100人分の分布・頻度情報を追加しました。

3) 【新規】経時変化情報を追加：

詳細二次調査参加者からご提供いただいた血漿試料に対してNMR/MSメタボローム解析を実施し、NMRメタボローム解析では1,000人分、43代謝物の定量値情報、MSメタボローム解析では代謝物キット解析を利用した約600人分、110代謝物の定量値情報を追加致しました。この拡張により加齢によるメタボロームの経時変化について、最大1,000人規模の比較解析を行うことが可能になります。また、約1,600人分の妊娠中あるいは産後の代謝物データも追加され、妊娠中及び産後1ヶ月の比較が可能となりました。

なお、いずれの代謝物についても、2018年拡張版jMorpの公開情報と同様に、代謝物間の相関情報、年齢層別の分布情報等を公開します。

今回のjMorp拡充は、宮城県で実施中のコホート調査の協力者の方々からご提供いただいた生体試料をもとに行われました。

■全ゲノムリファレンスパネルの拡充内容

これまで公開を行なってきた約 3,500 人からなる頻度パネル (3.5KJPNv2) の更新版として、4.7KJPN の構築を行いました。4.7KJPN は、東北メディカル・メガバンク計画による宮城県と岩手県でのコホート調査への協力者や、その他外部コホート事業における協力者、合計 4,773 人から構成されています。

表 2：拡充したコホート調査の内訳

コホート名	人数
東北メディカル・メガバンク計画による宮城県と岩手県でのコホート調査への協力者	4,378 (うち 993 人が新規追加)
独立行政法人国立病院機構長崎医療センターにおける協力者	188
ながはま 0 次予防コホート事業における協力者	40
国立がん研究センターにおける協力者	47 (新規追加)
J-MICC Study における協力者	60 (新規追加)
大阪大学眼科における協力者	30 (新規追加)
大阪大学ツインリサーチセンターにおける協力者	30 (新規追加)
合計	4,773

4.7KJPN はこれまでの 3.5KJPNv2 に含まれていた検体を引き継ぎつつ、新たに国立がん研究センターにおける協力者、J-MICC Study における協力者、大阪大学眼科における協力者、大阪大学ツインリサーチセンターにおける協力者等、東日本を中心にリクルートされた協力者合計 1,221 人のゲノム情報を追加することにより、より日本人として代表性の高いゲノム情報となりました。

4.7KJPN 頻度パネルに収録される SNV^{*13} (一塩基バリエント)・INDEL^{*14} (挿入・欠失) の数は以下の通りです。

表 3：4.7KJPN 頻度パネルに収録された SNV・INDEL 数

	SNV		INDEL	
	総数	新規数	総数	新規数
常染色体	59,199,967	24,949,665	8,238,823	4,543,503
X 染色体 (PAR1+PAR2)	2,316,891	1,051,758	341,289	191,811
X 染色体 (PAR1+PAR2+XTR)	2,318,572	1,052,789	341,547	191,915
ミトコンドリア	2,801	1,381		

- (dbSNP release 152 に記載のないものを新規として計算)
- (X 染色体は 2 種類の解析方法で解析された結果を公開。解析の詳細は Tadaka et al, 2019, *Human Genome Variation* をご参照ください)

また、これまで行なってきたアレル頻度情報以外に、新たにジェノタイプ頻度^{*15}情報の公開を始めました。これにより、より詳しい精度での解析が可能になります。ジェノタイプ頻度情報は、ORCID^{*16}と連携する認証を行い、データ移転契約（DTA：Data Transfer Agreement）をご確認いただくことで、jMorp ウェブサイトからダウンロード可能です。

【今後の展望】

当計画のコホート調査の詳細二次調査の参加者は現時点で既に数万人に及んでいます。今後、詳細二次調査参加者の検体のメタボローム解析を加速していくことで、男女、年代のバリエーションを持ち、かつ万人単位の経時変化を含む、世界でも類を見ないメタボローム情報のコレクションが構築され、ヒトの加齢変化を分子的、かつ網羅的に捉えることができる貴重な基盤となります。

東北メディカル・メガバンク計画では最終的に 8,000 人の解析データをもとにした全ゲノムリファレンスパネルの構築を目指しています。この全ゲノムリファレンスパネルが完成すれば、アレル頻度 0.1%程度の SNV データを格納できると考えており、かなり稀な疾患の原因解明まで含めて、当計画のリファレンスパネルが更に広く有効に使われることが期待されます。また、非常に解析が難しい Y 染色体の情報や、これまで収載している、SNV・INDEL 頻度のほかに、構造多型^{*17}の頻度情報を収載することも検討しており、「日本人全ゲノムリファレンスパネル」としてブラッシュアップし続けます。

【jMorp】

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

言語: 英語

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



【参考】

<東北メディカル・メガバンク計画について>

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と、個別化予防・医療の実現を目指しています。東北大学東北メディカル・メガバンク機構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、平成 25 年より合計 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等を実施して、試料・情報を収集したバイオバンク^{*18}を整備しています。東北メディカル・メガバンク計画は、

平成 27 年度より、国立研究開発法人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。

【用語解説】

- *1. メタボローム解析：生体内に含まれる代謝物（メタボライト）の網羅的な解析。
- *2. 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読により、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。
- *3. プロテオーム：生体内に含まれるタンパク質の網羅的な解析。
- *4. リファレンスパネル：ここでは疾患など健常時と異なる状態に対して、比較対照可能な参照用のデータ群を指す。
- *5. 詳細二次調査：当計画のコホート調査（*7 参照）において、参加者に対して最初に行った健康調査から約 4 年後に実施している 2 回目の詳細調査。数年おきに同じ人に対して健康調査を行うことにより調査開始時点からの前向きな経時的変化がわかる。
- *6. アレル頻度：ある集団における DNA バリエーションの塩基 (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）ごとに算出したもの。今回は対象となった日本人約 4,700 人中の頻度なので、最大で約 9,000 のうちにどれだけ検出されたか計算される。
- *7. コホート調査：特定の集団を一定期間追跡することによって、環境要因や遺伝的要因と疾病発生の関連を調べる調査。当計画の調査は「前向きコホート調査」である。
- *8. バイオマーカー：疾患の発症や進行を反映する生体内分子。
- *9. 日本人基準ゲノム配列：日本人のゲノム解読（次世代シーケンシング解析など）を行う際のひな型となるゲノム配列。ToMMo では 2019 年 2 月に「日本人基準ゲノム配列」初版 JG1 を発表している。
- *10. NMR メタボローム解析：NMR 解析は、生体分子を含む様々な分子を強力な磁場の中において、分子中の各原子が持つ核磁気モーメントを計測することにより、分子の構造や量を測定する解析方法。当計画では、NMR 解析を用いて生体中の代謝物を網羅的に解析（メタボローム解析）している。
- *11. MS メタボローム解析：MS（マススペクトロメトリー）は、物質を荷電粒子に変え、質量電荷比 (m/z) にて分離されたスペクトルとして検出する解析方法。生体内、食品及び環境に含有される様々な物質の存在量を測定することができる。当計画では、MS を用いて生体中の代謝物の定量的な解析を行っている。
- *12. 代謝物キット解析：当計画では、AbsoluteIDQ® p180 Kit (Biocrates 社製) による解析を行っており、これは代謝物を UHPLC-MS/MS を用いて一斉に定量分析できる試薬キットである。生体内アミン類に加え、多くの脂質代謝関連分子の検出が可能である。

- *13. 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。
- *14. 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。
- *15. ジェノタイプ頻度：遺伝子型頻度。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、対象となる約 4,700 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。
- *16. 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つくられた英数字コード。
- *17. 構造多型：ゲノム配列において、SNV や短鎖リードシーケンサーで検出できるような INDEL などの短い長さの多型ではなく、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。
- *18. 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。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集めた血液・尿などの生体試料だけではなく、それらを検査・解析した情報、調査票等から得られた情報も含む。

東北大学東北メディカル・メガバンク機構 <https://www.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
生命情報システム科学分野 教授 木下賢吾
電話番号：022-274-5952
生体分子解析分野 教授 小柴生造
電話番号：022-274-6016
ゲノム遺伝統計学分野 教授 田宮元
電話番号：022-274-5996

（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

（AMED 事業に関すること）

国立研究開発法人日本医療研究開発機構
基盤研究事業部 バイオバンク課
電話番号：03-6870-2228
E メール：tohoku-mm@amed.go.jp