



### 東北メディカル・メガバンク計画参加者における 先天性代謝異常症の原因バリエーションの保因者の頻度を推定 ～日本人一般集団における遺伝性疾患の理解へ向けて～

#### 【発表のポイント】

- 日本人の全ゲノムリファレンスパネル<sup>\*1</sup>を用いて、先天性代謝異常症<sup>\*2</sup>について原因遺伝子の病的バリエーションの保因者<sup>\*3</sup>の頻度を初めて推定し、ゲノムの多様性から劣性遺伝性疾患<sup>\*4</sup>の有病率の推定に取り組みました。
- 日本で新生児スクリーニング<sup>\*5</sup>対象となっている17の先天性代謝異常症の32の原因遺伝子について推定を行い、その結果、フェニルケトン尿症<sup>\*6</sup>やCPT2欠損症<sup>\*7</sup>において現実の有病率からの推定値と近い値を示し、ゲノムの多様性からの推定が可能であることを示唆しました。
- しかしながら、ゲノムおよび有病率からの保因者頻度の推定の比較結果は疾患ごとに異なり、病的バリエーションの発症への効果の程度が未解明であることや、検出されていない病的バリエーションが存在する、などの要因が示唆されました。
- 今後データの拡充やバリエーション解釈<sup>\*8</sup>法の進展などにより解明がさらに進むと、ゲノムのデータから一般集団での遺伝性疾患の潜在的な患者の割合の推定に役立つと期待されます。

#### 【概要】

遺伝性疾患は、形質のあらわれ方や家系での発症者のあらわれ方によって、何種類かに分類することができます。ヒトは、一人ひとりが両親からの2セットの遺伝情報を持っているので、疾患の原因遺伝子についても2つのコピーを持っています。優性(顕性)の遺伝性疾患は、疾患を起こす変異型を1コピーでも持っていれば、発症します。一方で劣性(潜性)の遺伝性疾患は、2コピー双方に機能欠損を起こす変異があり、その遺伝子の本来の機能を発揮するコピーが無いときに、疾患などの形質があらわれます。劣性の遺伝性疾患は、原因となる遺伝子の変異を2コピー双方に持つことで発症しますが、それを片方だけ持つ状態を「保因者」と呼びます。しかし、保因者の集団中での頻度が疾患ごとにどの程度であるのかは、はっきりとはわかっていませんでした。

疾患ごとに、発症者の頻度(有病率)が分かっている場合は、保因者の頻度は計算により推定可能ですが、実際に劣性の遺伝性疾患の原因変異を持つ人が一般集団中にどれくらいの割合で存在するかについては、これまで説明が困難でありました。

東北大学東北メディカル・メガバンク機構(ToMMo)では、日本人の全ゲノム解読を大規模に行い、全ゲノムリファレンスパネル(現在 3,552 人で、3.5KJPNv2 と呼ぶ)を構築して、ゲノムのバリエーションの頻度情報を国内外に公開しています。このパネルには、対象となった日本人において各バリエーションがどれだけの頻度で存在するかのデータが含まれています。本研究では、3.5KJPNv2 のデータを用いて、集団のゲノムの多様性は遺伝性疾患の有病率をどれくらい推定できるのか、という課題に取り組みました。具体的には劣性の遺伝性疾患である、先天性代謝異常症疾患の原因バリエーションの検出と保因者の頻度の推定を行い、各疾患の検査での陽性の人の割合に基づく既報の有病率との比較を行いました。

今回の研究で着目した遺伝性疾患は、フェニルケトン尿症など、原因遺伝子が分かっている 17 の先天性代謝異常症で、劣性(潜性)遺伝病です。これらの疾患は、治療法があり生化学的検査の精度が高いことなどから、新生児のスクリーニング対象として優先順位が高くなっています。そのため、多数の新生児が検査を受けており、疾患が見つかった割合について精度の高い統計データがあります。

本研究では、この 17 の疾患の原因として知られている 32 の遺伝子を解析対象としました。3.5KJPNv2 のバリエーション全体について生物・医学的アノテーション<sup>\*9</sup> とバリエーション解釈を行い、32 の遺伝子について、既報および推定された病的バリエーション検出を 4 種類の方法で行いました。検出された病的バリエーションとそれらの頻度を用いて、保因者の頻度の推定を行いました。一方で、過去の新生児スクリーニングで報告されている有病率から理論的に推定される保因者の頻度を得て、比較を行いました。4 種類の病的バリエーションの検出法に基づくゲノム変異の保因者の頻度を推定し、フェニルケトン尿症や CPT2 欠損症などの酵素欠損症では、有病率から推定された値と近い値を示しました。このことは、ゲノムデータより有病率を推定することが可能であることを示唆します。

しかしながら、ゲノムデータと有病率から推定された保因者頻度の比較結果は、疾患ごとに異なっていました。違いや推定値のずれを生んだ要因としては、ゲノムからの推定値の方が高ければ、バリエーションの発症への効果の程度が未解明であるなどの理由、一方でゲノムからの推定値の方が低ければ、病的バリエーションは他にも存在する、などの理由が考えられます。今後、更に研究を進めるにあたって、特に、1) 研究の進展に伴い知見が増えると病的バリエーションと判断されるものは増加する、2) 個々の病的バリエーションが特定の遺伝子型で実際に発病に至る割合については未解明であること、3) 頻度が高めのバリエーションであっても他の病的バリエーションとの共存状態で発病に貢献することがあり集団中の患者の割合に影響し得る、などに注意を払う必要があります。

本研究は、ヒト集団のゲノムの多様性が劣性の遺伝性疾患の有病率をどれくらい説

明できるのかについての初めての取組でしたが、今後、ゲノムデータと有病率それぞれに基づく保因者の頻度推定結果で違いが生じる方向と程度を明らかにしていくことで、ゲノムのバリエーション頻度と一般集団中の潜在的な患者数の割合との関係の理解が進むと期待できます。

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。本計画の事業の実施は、東北大学東北メディカル・メガバンク機構 (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 (IMM) とが連携して行っています。

## 【背景】

メンデル遺伝性疾患<sup>\*10</sup>は、遺伝様式ごとに特徴的な遺伝継承パターンを示し、多くは小児期に発症します。メンデル遺伝性疾患の一般集団での有病率は一般に低いのですが、劣性(潜性)遺伝病の遺伝的要因を、片方の染色体でのみ保有する保因者頻度の理論的推定値は、比較的高いと考えられています。メンデル遺伝病の原因遺伝子は既に同定されているものが多いのですが、一般集団中に存在する原因バリエントの種類やそれぞれの頻度については説明が不十分であり、日本民族集団特有のバリエントについてもよくわかっておりません。劣性遺伝病で原因遺伝子が1つの場合は、集団中の有病率を用いて、遺伝的な保因者頻度の理論的な推定値が得られてきました。しかし、劣性の遺伝性疾患の病的バリエントを保有する人が、実際に一般集団中にどれくらいの割合で存在するのか、ほとんどわかっていません。メンデル遺伝病の原因となるバリエント一つ一つは、一般集団内では非常に低い頻度で存在することが予想され、このような低頻度バリエントの存在実態の説明はこれまで困難でした。

ゲノムの解析技術の革新により、ヒトのゲノム全体をカバーする、全ゲノム解析や全エクソーム解析<sup>\*11</sup>が可能となり、ヒトゲノムの多様性の大量データが得られる時代となりました。また最近では、それぞれの国や地域で実施されるゲノムコホート研究<sup>\*12</sup>が増え、一人ひとりのための個別化医療・予防の基盤づくりと実現化を目的として推進されています。ToMMoにおいても、一般住民コホート参加者からの全ゲノム解読により、全ゲノムリファレンスパネルを構築し、研究コミュニティに提供し、利活用されています。このような活動から生産される集団のゲノムデータは、適切なアノテーションやバリエント解釈を通して、メンデル遺伝性疾患の原因バリエントの一般集団中での存在実態を調べることを可能にします。そこで私たちは、ヒト集団のゲノムの多様性は、遺伝性疾患の有病率をどれくらい推定できるのか、どのような関係になっているのか、という問いについて、取り組むことにしました。そして現状の全ゲノムリファレンスパネルの規模から、劣性の遺伝性疾患から研究することにしました。新生児スクリーニング対象の先天性代謝異常症は、酵素欠損症が多く劣性遺伝であること、そして多数の新生児が検査を受けるのでスクリーニング結果の有病率のデータがあることから、病的バリエントの集団中の頻度と現実の疾患の有病率との関係を調べるのに適していると考えました。そこで私たちは、3,552人の全ゲノム解読から構築された全ゲノムリファレンスパネル3.5KJPNv2を活用して、新生児スクリーニング対象の先天性代謝異常症の原因バリエントの同定と頻度の推定に取り組みました。

## 【方法・結果の詳細】

本研究で着目した17の疾患の原因として知られている32の対象遺伝子の領域内のバリエントを抽出しました。アノテーション情報やバリエント頻度情報を活用し、病的バリエントの同定と選抜を4種類(病的意義が確からしいバリエントのセットから、病的

意義の可能性の候補を含むセット)の方法で行いました。そして選抜された病的バリエーションのそれぞれの頻度から、集団内での頻度を推定し、病的バリエーションの保因者の頻度を推定しました。3,552 人の中で 17 の疾患の病的バリエーションのいずれかを持つ人の割合は、保守的な選抜方法では数%であり、既報のバリエーションをすべて含める方法では、10%を超えることが分かりました。一方で、新生児スクリーニングの過去の報告には、有病率の統計データがあります。そこで、原因遺伝子が 1 つである 11 の疾患について、有病率に基づく保因者の頻度を推定し、ゲノムデータからの推定値と比較しました。

フェニルケトン尿症は、アミノ酸代謝異常症の一つで、日本では約 6 万人に 1 人の割合で見つかります。原因遺伝子はフェニルアラニン水酸化酵素をコードする *PAH* 遺伝子ですが、病的バリエーションを 4 種類の方法(病的意義が確からしいバリエーションのセットから、病的意義の可能性の候補を含むセット)で選抜し、それぞれについてバリエーションの保因者の頻度を推定しました。4 つの推定値を有病率に基づく保因者の頻度と比較したところ、ClinVar<sup>\*13</sup> データベースに登録のある病的バリエーションおよびバリエーション解釈ガイドラインに基づく選択をしたとき、有病率に基づく保因者の頻度と近い値を示しました。

CPT2 欠損症(カルニチンパルミトイルトランスフェラーゼ-2 欠損症)は脂肪酸代謝異常症の一つで、幼児の突然死の原因となることから最近重要視されている遺伝性疾患の一つです。病的バリエーションの保因者頻度の 4 つの推定値を有病率に基づく保因者の頻度と比較したところ、ClinVar データベースに登録のある病的バリエーションおよびバリエーション解釈ガイドラインに基づく選択をしたとき、有病率に基づく保因者の頻度と近い値を示しました。

一方でシトリン欠損症<sup>\*14</sup> は、輸送体タンパク質(トランスポーター)の機能欠損によるのですが、原因遺伝子 *SLC25A13* の病的バリエーションから推定された保因者の頻度は、4 種類の方法においていずれも、有病率から推定された値を大きく上回っていました。このことは、国内の先行研究(\*14を参照)で報告されていた傾向と一致していました。

#### 【論文名】

タイトル: Estimating carrier frequencies of newborn screening disorders using a whole-genome reference panel of 3552 Japanese individuals

(日本語タイトル: 日本人 3,552 人の全ゲノムリファレンスパネルを用いた、新生児スクリーニング対象疾患の原因バリエーションのキャリア頻度の推定)

掲載誌: Human Genetics. 2019 Apr;138(4):389-409.

著者: Yumi Yamaguchi-Kabata, Jun Yasuda, Akira Uruno, Kazuro Shimokawa, Seizo Koshiba, Yoichi Suzuki, Nobuo Fuse, Hiroshi Kawame, Shu Tadaka, Masao Nagasaki, Kaname Kojima, Fumiki Katsuoka, Kazuki Kumada, Osamu Tanabe, Gen Tamiya, Nobuo Yaegashi, the Tohoku

【用語解説】

- \*1 全ゲノムリファレンスパネル 3.5KJPNv2：東北メディカル・メガバンク計画で実施された、一般の日本人 3,552 人の全ゲノム解読により、検出されたゲノムのバリエーションから構築された日本人ゲノムの変異パネル。バリエーションの頻度情報は、jMorp (<https://jmorp.megabank.tohoku.ac.jp/>) より公開されている。<https://www.megabank.tohoku.ac.jp/news/27839>
- \*2 先天性代謝異常症：生まれつきの遺伝的な要因により、体の構成成分の新陳代謝やエネルギー産生の代謝の一部が機能しなくなっていることによって、発症する疾患。遺伝性の代謝機能の欠損の多くは、代謝に関わる特定の酵素や分子をコードする遺伝子に変異があることで機能欠損型となっている。
- \*3 保因者（キャリア）：ここでは常染色体劣性（潜性）の遺伝性疾患において、病気の原因となる遺伝的要因を一つ、片方の染色体でのみ、ヘテロの状態で保有している人のこと。もう片方の染色上の該当遺伝子の産物が正常な機能を発揮していれば、発症しない。
- \*4 劣性（潜性）遺伝性疾患：メンデル遺伝性疾患の特定の原因遺伝子において、父方由来と母方由来の両方（2 コピー）が、ともに機能欠損のタイプであり、正常の機能を発揮するコピーを持たないときに、発症する遺伝性疾患。家系では、世代をスキップして発病者が現れる。直近の遺伝学用語改訂（日本遺伝学会）では、検討を経て、劣性を潜性へ（そして優性を顕性へ）と改訂している。
- \*5 新生児スクリーニング：先天性代謝疾患や先天性内分泌疾患の早期発見および早期治療を可能にするため、新生児を対象に、生まれて数日後の血液を用いて行われる検査。歴史的には、フェニルケトン尿症の検出から始まり、生物学的な手法により検査が行われてきた。最近では、質量分析法を活用したタンデム・マス試験が導入されており、多種類の代謝物を短時間に測定できるようになった。日本でのタンデム・マス試験一次対象疾患の 16 疾患は、治療法が確立しており、かつ検査の精度が高い、などの理由で選ばれている。二次対象疾患は、検査対象としての追加が検討されているものであり、CPT2 欠損症などが含まれる。
- \*6 フェニルケトン尿症：アミノ酸代謝異常症の一つ。フェニルアラニンをチロシンに変換する酵素（フェニルアラニン水酸化酵素）の機能欠損のため、

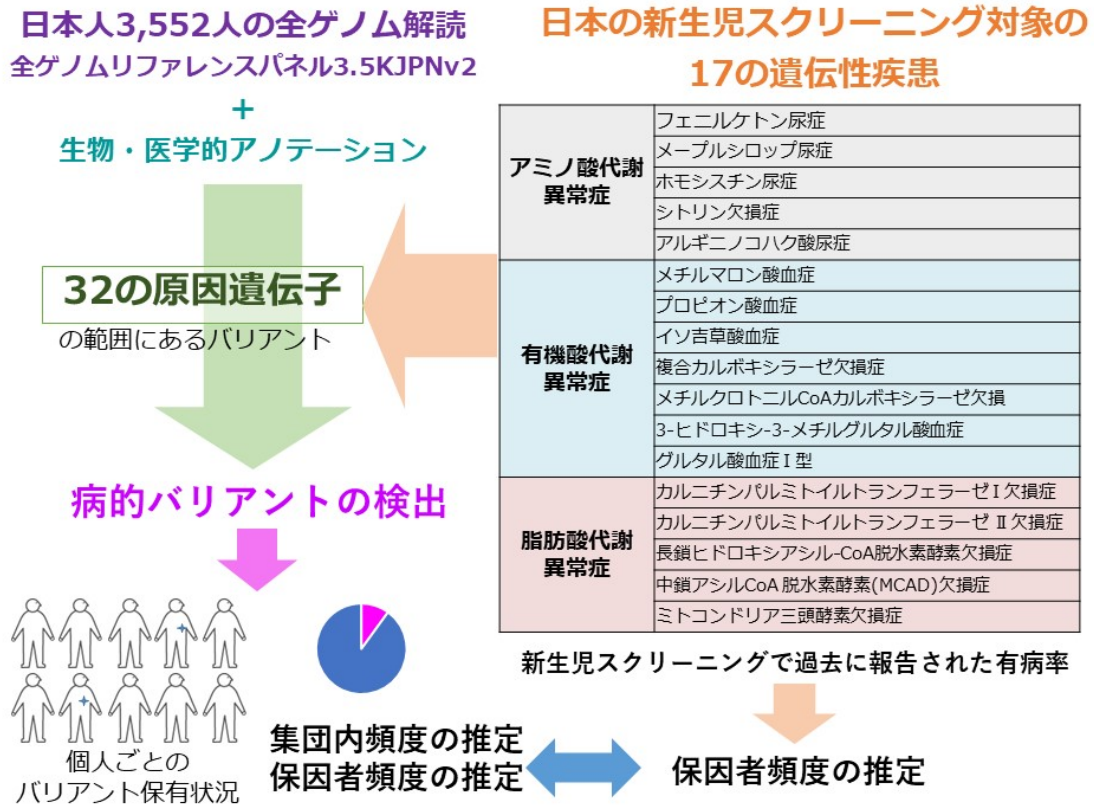
身体にフェニルアラニンが蓄積し、チロシンが少なくなる。血中のフェニルアラニンの濃度が高い状態が続くと、発達遅滞などの神経症状を示す。チロシンの欠乏により、色素産生に影響するので、髪や皮膚の色が薄くなる。過剰なフェニルアラニンは、フェニルケトン体として尿中に排泄されるため、フェニルケトン尿症として知られている。

- \*7 CPT2 欠損症：カルニチンパルミトイルトランスフェラーゼ-2 欠損症。体内で脂肪を分解できないために、エネルギーとして活用することができない、脂肪酸代謝異常症の一つ。エネルギー産生不全に陥ることが、重篤な低血糖症状につながる。そしてエネルギーを必要とする脳や筋肉に影響するため、意識障害、痙攣、などの症状を示し、時には突然死に至ることもある。新生児スクリーニングでは二次対象疾患と位置づけられてきたが、最近では重要視されている。
- \*8 バリエーション解釈： 遺伝子のバリエーションが病因性を持つかどうか（病的バリエーションであるかどうか）についての解釈を行うこと。2015年にアメリカ臨床遺伝・ゲノム学会(ACMG)と Association for Molecular Pathology (AMP) は、ゲノムのバリエーション解釈のガイドラインを共同で発表(Richards et al. Genet Med 17: 405-24, 2015)しており、このガイドラインは、現在多くの専門家と研究者に使用されている。
- \*9 アノテーション： ここでは、ゲノムの塩基配列やバリエーションについて、生物・医学的な注釈付けをすること。
- \*10 メンデル遺伝性疾患： 1つの原因遺伝子の変異によって生じる遺伝性疾患で、その遺伝継承はメンデルが発見した遺伝の法則に従う。
- \*11 全エクソーム解析：ゲノムのうち、タンパク質に翻訳される部分（エクソンと呼ぶ）を全て解析する手法のこと。
- \*12 ゲノムコホート研究：ヒトの一定の集団を観察し、疾患の発症原因などを調べる研究をコホート研究と呼ぶが、そのうち、特に、対象となる集団からゲノム情報を収集して行う研究を呼ぶ。
- \*13 ClinVar：米国NCBI（国立生物工学情報センター、国立衛生研究所（NIH）の一部門）が運用するデータベースで、医学的に重要なヒトの変異の情報を疾患などの表現型の情報とともに収載している。
- \*14 シトリン欠損症：シトリンとは、肝臓細胞のミトコンドリア膜に存在するアスパラギン酸・グルタミン酸の輸送体タンパク質（トランスポーター）であり、その働きは細胞の中で、糖・アミノ酸・脂質・エネルギー代謝に関係している。新生児スクリーニングでは、二次対象疾患に指定されている。シトリンの機能低下により細胞質内NADHの蓄積が起これ、糖質を過剰に摂取すると、代謝不全を起こす。東アジア～東南アジアで比較的頻度の高い疾患。新生児や乳児期では新生児肝内胆汁うっ滞症、成人期では、成人発症 II 型シトルリン血症、のように年齢によって異なる症状を示す。

国内の先行研究により、*SLC25A13* 遺伝子に病的バリエントをヘテロで持つ人の割合は高い(65 人に 1 人)が、期待されるホモ接合頻度に対して、実際に発症する人の割合は低いことが報告されており、ホモ接合体であっても症状が出ない場合があることが推測されている (Kobayashi et al. 2003 Mol Genet Metab 80: 356-9; Tabata et al. 2008 J Hum Genet 53: 534-45; Kikuchi et al. 2012 Mol Genet Metab 105: 553-8 ) 。

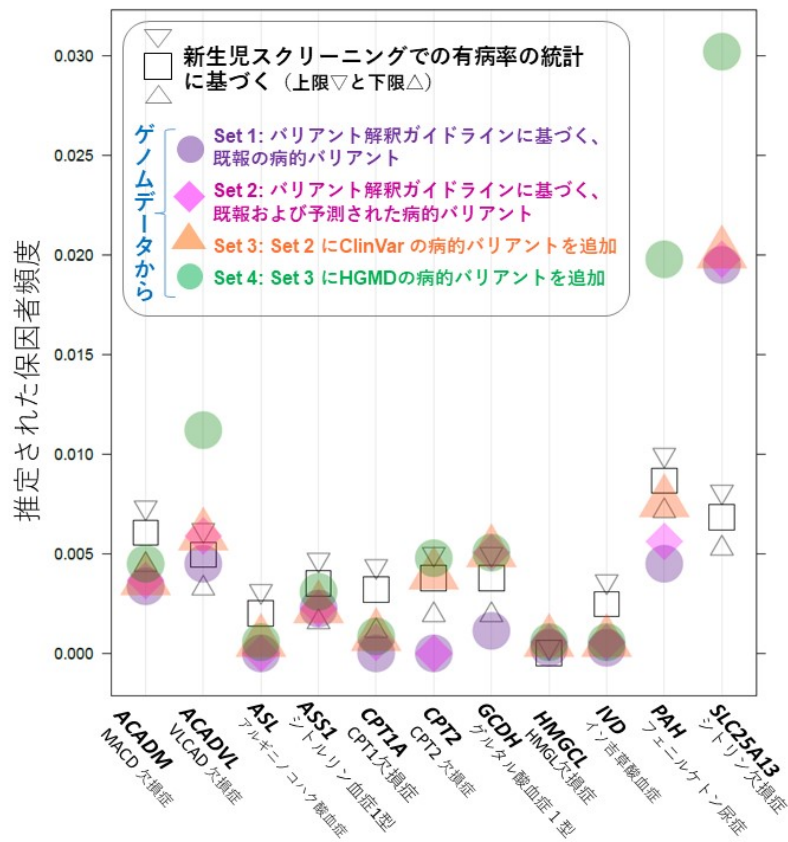


【図 1: 本研究の解析の概要】



全ゲノムリファレンスパネル 3.5KJPNv2 のバリエーション全体について、医学・生物学的アノテーションを行い、日本の新生児スクリーニング対象疾患の原因となる 32 の遺伝子を対象に解析を行った。アノテーション情報を活用したバリエーション解釈により、病的バリエーションの抽出を 4 種類の方法（病的意義が確からしいバリエーションのセットから、病的意義の可能性の候補を含むセット）で行った。検出された病的バリエーションから、集団内の頻度、および保因者頻度の推定を行った。

【図 2: 推定された保因者頻度の比較】



原因遺伝子が 1 つである 11 の疾患について、過去の新生児スクリーニングで報告されている有病率から推定された保因者頻度と、今回の研究でゲノムのバリエーションデータの解析から推定された保因者頻度の比較を行った。

【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  
ゲノム解析部門 助教

山口 由美(やまぐち ゆみ)

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:f-nagami@med.tohoku.ac.jp