

国立健康危機管理研究機構 国立国際医療研究センター

九州大学

東京大学

情報・システム研究機構 データサイエンス共同利用基盤施設

ライフサイエンス統合データベースセンター

情報・システム研究機構 国立遺伝学研究所 生命情報・DDBJ センター

日本人とサウジアラビア人のゲノム情報を反映した 新しい「ゲノム地図」を作成

【発表のポイント】

- 10 人の日本人と 9 人のサウジアラビア人の実質的に完全なゲノム配列を決定
- 個人間のゲノム配列の違いを表す「パンゲノムグラフ」によって複雑なゲノム構造が明らかに
- ロングリードシーケンサーなどの最先端の技術を活用
- 地域に特化した精密医療・遺伝子診断の基盤を提供

【概要】

国立健康危機管理研究機構（東京都新宿区、理事長：国土典宏） 国立国際医療研究センター ゲノム医科学プロジェクト 河合洋介 副プロジェクト長、サイデ・アシュリ特任研究員、徳永勝士プロジェクト長らは、アブドラ国王科学技術大学（サウジアラビア王国）とライフサイエンス統合データベースセンター、九州大学、東京大学、国立遺伝学研究所と共同で、日本人とサウジアラビア人の完全長ゲノム配列を決定して、公共データベースに登録しました。

これまでのゲノム研究では、特定の 1 人のゲノム配列を基準とする方法が使われてきましたが、これでは地域に特有の遺伝的な違いが反映されにくいという課題がありました。今回、日本とサウジアラビアの研究者を中心とする国際チームで、両国のゲノム多様性を反映した新しいパンゲノムグラフ（注 1）「JaSaPaGe」を開発しました。このパンゲノムグラフは、日本人 10 人とサウジアラビア人 9 人の高品質な全ゲノムデータをもとに構築されたもので、従来の方法に比べてより多くの遺伝的変異を高精度に検出可能であることが示されました。研究成果は FAIR 原則（注 2）に基づき、誰でも自由に利用可能な形でデータを公開しています。これにより、日本や中東地域を含む多様な人々の遺伝情報を反映した、より正確で公平なゲノム研究が進展することが期待されます。

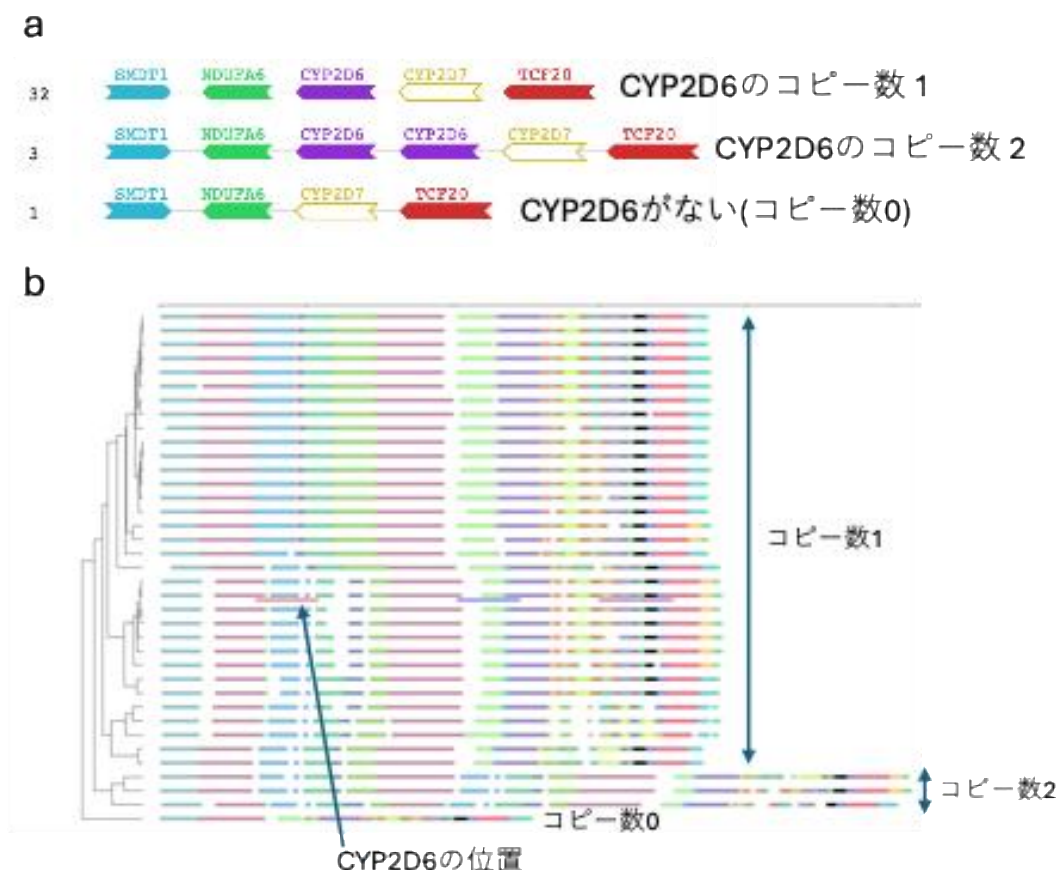
【研究の背景】

ヒトゲノム解析では、これまで単一の個人のゲノムに相当する配列が参照配列として使われており、個人差の大きい構造的変異や集団固有の遺伝情報が反映されにくいという課題がありました。

近年、新たなゲノム多様性の表現方法として「パンゲノムグラフ」が登場し、複数人のゲノム配列を統合的に扱うことで、より多様な遺伝的変化を捉えられるようになり、国際的にパンゲノムグラフを作る取り組みが進められていますが、こうした国際的な取り組みでは日本人やアラブ人のサンプルはほとんど含まれておらず、世界の約 6 億人に相当する集団が“ゲノム地図”から漏れていたのが現状でした。この背景を受け、本研究では日本人とサウジアラビア人のゲノム多様性を反映した独自のパンゲノムグラフ（JaSaPaGe）を構築することで、これらの地域に根ざしたより正確なゲノム解析の基盤を提供することを目的としました。

【研究結果】

本研究チームは、日本人 10 人、サウジアラビア人 9 人（うち 1 人は既存データ）から得た高品質なゲノムデータをもとに、2 組の染色体に由来する配列を含む「二倍体ゲノムアセンブリ」（注 3）を行い、それを基に日本人の配列・サウジアラビア人の配列・従来の参照配列（GRCh38, CHM13-T2T）からなるパンゲノムグラフを作成しました。これらの新しい参照ゲノムを使って短鎖リード配列からの変異検出（バリエーションコール）を行った結果、従来の参照配列を使うよりも約 3～10% 多くの遺伝的変異が検出可能であることがわかりました。このことは対象集団に特化したパンゲノムグラフを使用することにより変異の同定精度が大幅に向上することを意味しています。また、薬物代謝に関わる重要な遺伝子 CYP2D6 のコピー数（CNV）において、集団ごとに異なる構造的多型が明瞭に識別可能であることも実証され、個別化医療（プレシジョン・メディシン）への応用の可能性を示しました（下図）。



薬剤代謝遺伝子CYP2D6のコピー数の個人間の違い

- ゲノム配列から同定したCYP2D6周辺の遺伝子の並び。CYP2D6を1コピーもつ配列が最も多いが、2コピーやCYP2D6を持たない配列も見つかりました。
- この領域の配列を似た塩基配列の領域で色分けした図。コピー数が2の配列で周辺の配列も含めて重複していることがわかりました。

本研究で得られたすべてのゲノムデータ・アセンブリ・解析コード・パンゲノムグラフは、FAIR原則に則ってオープンアクセスで公開されています。データは日本 DNA データバンク（DDBJ）に登録されており、誰でも自由に利用・再解析が可能です（BioProject ID：PRJDB19680）。このように、本研究は単にゲノム解析の成果にとどまらず、国際的なデータ共有・再利用の模範モデルとしても高く評価される内容となっています。

これらの成果により、日本・サウジアラビアなど地域に特化した精密医療・遺伝子診断の基盤が大きく前進するとともに、国際的なパンゲノムグラフを構築する取り組みにも重要な貢献を果たすと期待されます。

共同責任著者は本研究の意義について以下のように述べています。

「私たちの推定では、適用対象集団と整合しない基準を使用しているため、遺伝性疾患の患者の最大 12% が診断を受けることができていません。私たちが構築したパンゲノムグラフによってバリエーションの検出精度が改善し、この問題を解決することができます」 - マラック・アベド・アルタガフィ（エモリー大学医学部教授）

「日本とサウジアラビアは、アジアのほぼ反対側に位置し、長い間隔てられてきました。これらの大きく異なる集団を比較することにより、集団特異的なパンゲノムグラフの有効性を検証するのに役立ちました」 - ロバート・ヘーンドルフ（アブドラ国王科学技術大学准教授）

【発表者・研究者等情報】

国立健康危機管理研究機構 国立国際医療研究センター ゲノム医科学プロジェクト 副プロジェクト長
情報・システム研究機構 国立遺伝学研究所 生命情報・DDBJ センター 特命教授
河合 洋介

国立健康危機管理研究機構 国立国際医療研究センター ゲノム医科学プロジェクト 特任研究員
Saeideh Ashouri（サイデ・アシュリ）

国立健康危機管理研究機構 国立国際医療研究センター ゲノム医科学プロジェクト プロジェクト長
徳永 勝士

九州大学 生体防御医学研究所 教授
長崎 正朗

九州大学 生体防御医学研究所 教授
大川 恭行

東京大学 大学院新領域創成科学研究科 教授
鈴木 穰

情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター 特任教授
片山 俊明

【論文情報】

雑誌名：Scientific Data

題 名：Phased genome assemblies and pangenome graphs of human populations of Japan and Saudi Arabia

著者名：Maxat Kulmanov*, Saeideh Ashouri*, Yang Liu, Marwa Abdelhakim, Ebtehal Alsolme, Masao Nagasaki, Yasuyuki Ohkawa, Yutaka Suzuki, Rund Tawfiq, Katsushi Tokunaga, Toshiaki Katayama, Malak S Abedalthagafi#, Robert Hoehndorf#, Yosuke Kawai#

(*:筆頭著者, #:責任著者)

DOI: 10.1038/s41597-025-05652-y

URL: <https://www.nature.com/articles/s41597-025-05652-y>

【研究助成】

研究は以下の研究資金による支援のもと行われました。

科学技術振興事業団（JST）ライフサイエンスデータベース統合推進事業・統合化推進プログラム (DICP):JPMJND2302

日本医療研究開発機構（AMED）：JP21wm0425009, JP22fk0210111, JP22tm0424222, JP23ek0109675, JP23ek0109672, JP23fk0210138, JP23ek0210194, JP24gm2010001

日本学術振興会（JSPS）科学研究費補助金：JP18H05527, JP21H02681, JP24H02323

【用語解説】

注1 パンゲノムグラフ

ヒトのゲノムは、全体としては共通した配列を持っていますが、人によって少しずつ違いがあります。これまでの遺伝子解析では、ひとつの代表的なゲノム配列（参照配列）だけを基準に使っていましたが、それでは他の人にしかない遺伝的な特徴が見つかりにくいという問題がありました。パンゲノムグラフは、複数の人のゲノム情報をまとめて比較・統合できる、新しいデータの表現方法です。個人ごとの違いもすべて含めて記録するため、これまで見つけにくかった変異や特徴も検出しやすくなります。

注2 FAIR 原則

FAIR 原則とは、研究で得られたデータをより多くの人にとって使いやすくするための国際的な考え方です。FAIR という言葉は、「見つけやすく (Findable)」「アクセスしやすく (Accessible)」「他のシステムと連携しやすく (Interoperable)」「再利用しやすい (Reusable)」という 4 つの基本的な考えを表しています。この原則に従ってデータを整理・公開することで、研究者だけでなく、医療関係者や開発者などさまざまな人が必要なデータをすぐに見つけて使えるようになります。また、他の研究との組み合わせや再利用もスムーズになり、科学の発展や社会への応用が加速します。

注3 二倍体ゲノムアセンブリ

ヒトのゲノム（遺伝情報の全体）は、父親と母親から 1 セットずつ、合計 2 セットの DNA で構成されています。これを「二倍体（にばいたい）」と呼びます。「二倍体ゲノムアセンブリ」とは、この 2 つの DNA セットをそれぞれ分けて、正確に再構築することを意味します。つまり、父由来と母由来の違いをきちんと分けて読み解くことで、より詳細で正確な遺伝情報が得られます。

【問い合わせ先】

《研究に関すること》

国立健康危機管理研究機構 国立国際医療研究センター ゲノム医科学プロジェクト

河合 洋介

Tel : 03-3202-7181

《取材に関すること》

国立健康危機管理研究機構 危機管理・運営局 広報管理部

Tel : 03-3202-7181 E-mail : press@jihs.go.jp

九州大学 広報課

Tel : 092-802-2130 E-mail : koho@jimukyushu-u.ac.jp

東京大学 大学院新領域創成科学研究科 広報室

Tel : 04-7136-5450 E-mail : press@k.u-tokyo.ac.jp

情報・システム研究機構 データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター 広報担当
E-mail : public_relations@dbcls.rois.ac.jp

情報・システム研究機構 国立遺伝学研究所 生命情報・DDBJ センター 広報室
E-mail : prkono@nig.ac.jp