PRESS RELEASE

東京大学

病気に関わる“ゲノムの迷宮”を解く

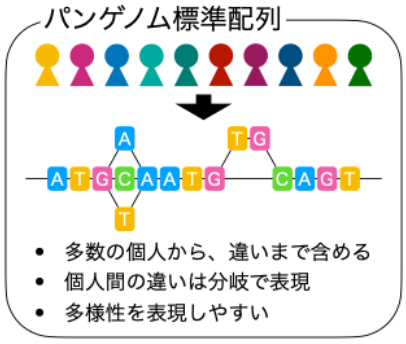
――免疫・疾患に関わる遺伝子領域を日本人集団で高精度に解読――

発表のポイント

◆複数の DNA 解読技術を統合し、日本人 10 人のゲノム配列を極めて高い精度で決定することで、日本人集団の「パンゲノム配列」を構築しました。

◆免疫・疾患に関わる複雑な遺伝子領域の完全解読率を従来の 46.8%から 91.2%へ高め、新たなゲノム配列構造や、進化の過程における偏りを示唆する変異パターンを明らかにしました。

◆本成果は、日本人集団のゲノム多様性をより精密に理解する基盤となり、将来の疾患研究や個別化医療への貢献が期待されます。



パンゲノム標準配列の概念図

概要

東京大学大学院新領域創成科学研究科の鈴木慶彦特任助教（研究当時）と森下真一教授（研究当時）、カリフォルニア大学サンタクルーズ校の Benedict Paten 教授および Karen Miga 准教授らによる研究グループは、日本人 10 人のゲノム配列を極めて高い精度で決定し、免疫・疾患に関わる複雑な遺伝子領域における完全なゲノム配列と遺伝子構造の多様性を明らかにしました。

ヒトゲノムは一人ひとり少しずつ異なります。そこで、複数の人のゲノム配列を合わせてヒトゲノムの「標準配列」を個人差の多様性を表現しながら拡張するパンゲノム（注 1）研究が近年進んでいます（図 1）。しかし免疫や病気に関わる遺伝子が多く集まる、似た塩基配列が繰り返し並んだゲノム領域は、従来の技術では読み切ることが困難で大きな課題でした。研究チームは 10 人の日本人男性由来サンプルから、DNA の塩基配列を長く正確に読む複数の技術を組み合わせ、20 個のほぼ完全なハプロタイプ（注 2）ごとのゲノム配列を構築しました。その結果、医学・生物学的に重要な 30 の代表的な遺伝子領域で、完全に隙間なく塩基配列を再構築できた割合を既存研究の 46.8%から 91.2%へ高めました。それによって免疫や疾患に関わる重要な遺伝子領域で新規のゲノム構造を発見しました。本成果は、日本人集団のゲノム多様性のより正確な理解に役立ち、将来の疾患研究や個別化医療の基盤を強めることが期待されます。

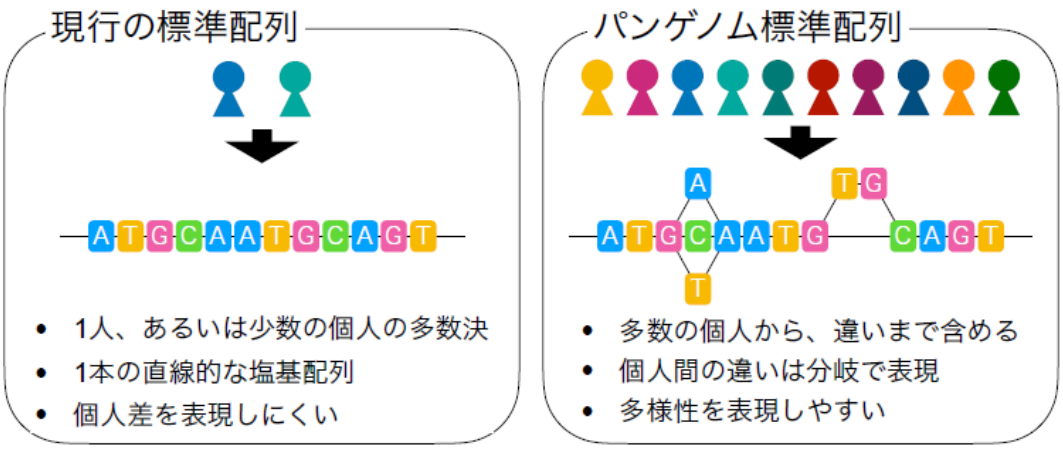


図 1：パンゲノム標準配列の基本的なアイデア

これまで、ゲノム解析には限られた個人の配列を代表として直線状に構築された標準配列（図左）が主に用いられている。パンゲノム標準配列（図右）では、個人ごとの塩基配列の違いを取り込むことで標準配列を拡張し、多様性を表現・解析しやすくすることを目指す。

発表内容

〈研究の背景〉

ヒトゲノムの標準配列は研究や医療の基盤ですが、現在最も利用されている標準配列は限られた個人の配列を代表として構築されているため、集団ごとの個人差の多様性を十分に表せないことがありました。Human Pangenome Reference Consortium（HPRC）などの国際コンソーシアムは、複数の人のゲノムを束ねた「パンゲノム標準配列」を整備してきましたが、初期の HPRC パンゲノムデータには日本人サンプルが含まれていませんでした。さらに、*KIR*、*SMN*、*DEFB*（注 3）など疾患に関係する遺伝子領域の多くは、よく似た配列が長い領域にわたって繰り返す「セグメント重複」（注 4）という構造を含み、パンゲノム標準配列においても塩基配列が途切れたり誤って構築されたりしやすいことが大きな課題でした。

〈研究の内容〉

本研究では、日本人として収集された 10 人の健康な男性由来細胞培養株から超長鎖 DNA シーケンシング技術（注 5）を含めた 3 つの異なる種類の DNA 解読技術を用いて DNA 配列を取得しました。これら 3 種類の配列を統合し、染色体規模で 20 個のほぼ完全なハプロタイプの塩基配列を構築しました。得られた塩基配列は非常に高精度で、すべての人についてゲノム配列の半分以上を連続して 1 億塩基対以上途切れ目なくつなげることができ、医学・生物学的に重要な 30 の代表的な遺伝子領域では平均 91.2%のハプロタイプで塩基配列を完全に再構築できました（図 2A）。*KIR* 遺伝子領域や *SMN* 遺伝子領域では従来のパンゲノムでは検出されていなかったマイナーな配列を発見しました（図 2B、C）。さらに、*SMN* 遺伝子や *DEFB* 遺伝子の周辺領域からは非対称的な偏った遺伝子変換（注 6）によるものと推測される塩基配列の変異パターンを検出し、これらのパターンは非ランダムな進化過程の結果である可能性を見出しました。本研究の日本人パンゲノムグラフは Zenodo（https://doi.org/10.5281/zenodo.14160240）で公開されています。また、パンゲノムグラフ構築に使用されたハプロタイプ別ゲノム配列やシーケンシングデータは HPRC のウェブサイト（https://humanpangenome.org/hprc-data-

release-2/）および NCBI データベース（アクセッション番号：PRJNA1184421-PRJNA1184440 および PRJNA1459655）で公開されています。

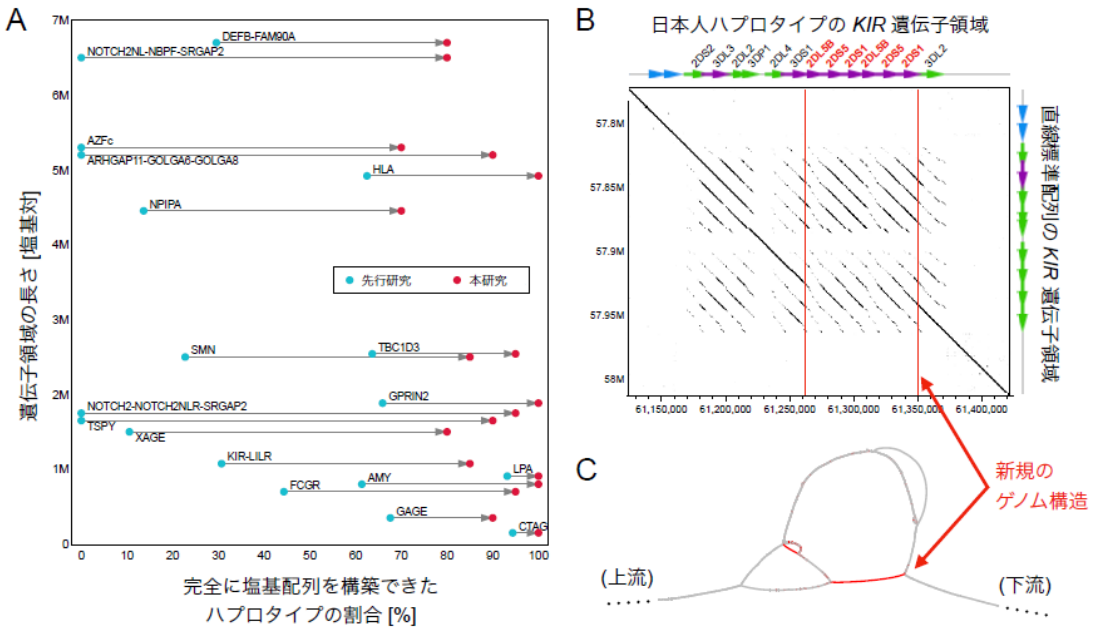


図 2：主要な遺伝子領域における結果

（A）代表的ないくつかの複雑な遺伝子領域について、先行研究（HPRC Year-1）と本研究の日本人ゲノム配列で、ハプロタイプを完全に再構築できた割合を比較した。本研究では、多くの領域で完全再構築率が大きく向上した。

（B）19 番染色体の *KIR* 遺伝子領域について、本研究で得られた 1 つの日本人ハプロタイプと直線標準配列を比較した図。上と右にある矢印は各配列上の遺伝子を表し、そのうち赤字は本研究で新たに見つかった *KIR* 遺伝子の並びである。黒枠内では、2 つのゲノム配列間で類似する領域の対応関係が斜線で表示され、濃い線ほど配列類似度が高いことを示す。赤点線で囲まれた領域の外側では、長く連続した斜線が見られ、両配列が構造的に対応していることが分かる。一方、赤点線内では、*KIR* 遺伝子間の類似性に由来する短い平行な斜線は見られるものの、領域全体を通した長い一対一の対応は認められない。このことは、この日本人ハプロタイプが直線標準配列とは大きく異なる構造であることを示している。

（C）*KIR* 遺伝子領域の日本人パンゲノムグラフ。B で示した日本人ハプロタイプに特有の配列を赤色で、その他の共有配列を灰色で示してある。B の新規ゲノム構造は、主に中央下部の赤色の経路に対応する。

〈今後の展望〉

今後は HPRC などの国際コンソーシアムと協力し、日本人の高精度ゲノム配列を世界中の他の多様な集団のデータと組み合わせ、次世代の高精度パンゲノム標準配列の整備を目指します。それによって、従来の限られた個人の標準配列では見落とされていた構造変異や遺伝子コピー数の違いを調べやすくなります。将来的には、このパンゲノム標準配列は進化や疾患の研究、ひいては個別化医療のための AI 開発の重要な基盤となることが期待されます。

発表者

東京大学大学院新領域創成科学研究科
森下　真一　研究当時：教授
現：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　バイオ生成 AI 研究開発センター　特任教授
鈴木　慶彦　研究当時：特任助教
現：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　バイオ生成 AI 研究開発センター　特任講師

カリフォルニア大学サンタクルーズ校
Benedict Paten　教授
Karen Miga　准教授

論文情報

雑誌名：Nature Communications
題　名：Accessing medically relevant complex regions with a pangenome graph of 20 near-complete Japanese haplotypes
著者名：Yoshihiko Suzuki*, Chie Owa, Haruka Kobayashi, Ryo Nakabayashi, Brandy McNulty, Ivo Violich, Benedict Paten, Karen H. Miga, Shinichi Morishita*
DOI：10.1038/s41467-026-73461-x
URL：https://www.doi.org/10.1038/s41467-026-73461-x

研究助成

本研究は、日本医療研究開発機構（AMED）のゲノム医療実現バイオバンク利活用プログラム（ゲノム医療実現推進プラットフォーム・先端ゲノム研究開発）「ヒトゲノム De Novo 情報解析テクノロジーの創出」、JSPS 科研費 JP22H04925（PAGS）および JP24K18091、NIH/NHGRI R01HG011274 および UM1HG010971 の支援により実施されました。また、東京大学情報基盤センターのスーパーコンピュータ（mdx）の支援を受けました。

用語解説

（注 1）パンゲノム
複数の人のゲノム情報をまとめ、1 本の標準配列だけでは表現しきれない個人ごとのゲノム配列の違いも含んだ「ゲノムの詳細地図」。英語だと pangenome で、“pan-”は「全ての」という意味。パンゲノムを使うことで、個人差や集団差を表現・解析しやすくなる。

（注 2）ハプロタイプ
父親または母親から受け継いだ 1 組のゲノム配列。ヒトは通常、父親由来と母親由来の 2 つのハプロタイプを持つ。本研究では、個人ごとに 2 つのハプロタイプを区別して塩基配列を組み立てた。

（注 3）*KIR*、*SMN*、*DEFB*
KIR は免疫細胞の働きに関わる受容体遺伝子群。*SMN* は脊髄性筋萎縮症（SMA）と関係する遺伝子群で、コピー数や周辺の遺伝子との組み合わせが症状理解の手がかりになる可能性が示唆さ

れている。*DEFB* は、細菌などから体を守る抗菌ペプチドである β ディフェンシンに関わる遺伝子群で、個人ごとの *DEFB* 遺伝子の数の違いが乾癬やクローン病などとの関連で研究されている。

（注 4）セグメント重複
ゲノム中で長くよく似た塩基配列が複数コピー存在する領域。一般的に 1,000 塩基対以上で 90%以上の類似度をもつ領域を指す。免疫や疾患に関わる遺伝子を含むことがある。塩基配列が似ているため区別しながら配列を決定することが難しい。

（注 5）超長鎖 DNA シーケンシング技術
Oxford Nanopore Technologies 社による、最大 100 万塩基対程度までの極めて長い DNA 分子を一度に読み取ることができる DNA 解読技術。

（注 6）遺伝子変換
よく似た DNA 配列の間で、一方の配列の一部がもう一方の配列を鋳型として書き換わる現象。これにより、ある場所の塩基配列の変異が、別のよく似た場所にコピーされたように見えることがある。

問合せ先

〈研究に関する問合せ〉
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
バイオ生成 AI 研究開発センター
特任教授　森下　真一（もりした　しんいち）
E-mail：moris@edu.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
バイオ生成 AI 研究開発センター
特任講師　鈴木　慶彦（すずき　よしひこ）
E-mail：suzuki.yoshihiko@rois.ac.jp

〈報道に関する問合せ〉
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