

PRESS RELEASE

2024年4月17日
理化学研究所
静岡県立総合病院
静岡県立大学
東京大学医科学研究所全ゲノム解析で明らかになる日本人の遺伝的起源と特徴
—ネアンデルタール人・デニソワ人の遺伝子混入と自然選択—

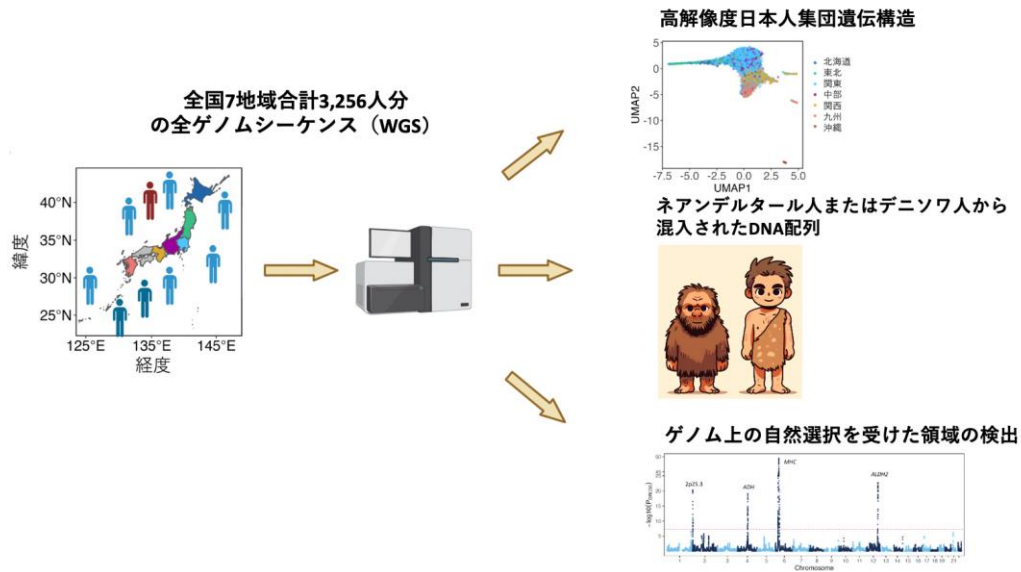
概要

理化学研究所（理研）生命医科学研究センターゲノム解析応用研究チームの寺尾知可史チームリーダー（静岡県立総合病院臨床研究部免疫研究部長、静岡県立大学薬学部ゲノム病態解析講座特任教授）、劉暁溪上級研究員（研究当時：ゲノム解析応用研究チーム研究員；静岡県立総合病院臨床研究部研究員）、東京大学医科学研究所附属ヒトゲノム解析センターシーケンス技術開発分野の松田浩一特任教授らの共同研究グループは、大規模な日本人の全ゲノムシーケンス（WGS）^[1]情報を分析し、日本人集団の遺伝的構造、ネアンデルタール人^[2]およびデニソワ人^[3]由来のDNAと病気の関連性、そしてゲノムの自然選択が影響を及ぼしている領域を複数発見しました。

本研究成果は、日本人集団の遺伝的特徴や起源の理解、さらには個別化医療^[4]や創薬研究への貢献が期待されます。

今回、共同研究グループは、バイオバンク・ジャパン（BBJ）^[5]が提供した3,256人分の日本人の全ゲノム情報を分析しました。この研究を通じて、日本人の祖先に関わる三つの源流（縄文系祖先、関西系祖先、東北系祖先）の起源を明らかにしました。また、現生人類（ホモ・サピエンス）の最も近縁とされる古代型人類ネアンデルタール人やデニソワ人から受け継いだ遺伝子領域を特定しました。一部の日本人が持つNKX6-1遺伝子領域は、2型糖尿病リスクと関連しており、この領域がデニソワ人由来であることが明らかになりました。さらに、日本人の遺伝子における自然選択が作用している領域も同定しました。

本研究は、科学雑誌『Science Advances』オンライン版（4月17日付：日本時間4月18日）に掲載されます。



大規模な日本人の全ゲノムシーケンスから日本人集団の遺伝的構造などを解明

背景

全ゲノムシーケンス (WGS) データは、遺伝学および生物医学で不可欠なものとなっています。WGS データの分析を通じて、人間の遺伝的多様性や自然選択のプロセスを理解する上での新たな扉が開かれました。また、個別化医療や創薬研究におけるゲノムワイド関連解析 (GWAS) の解析能力を格段に向上させ、病気に関わるバリエーションの同定や治療標的の特定へとつながっています。

しかし、日本人集団を対象とした WGS 研究は、限定的な規模にとどまってきました。このギャップを埋めるため、共同研究グループは全国 7 地域から集めた 3,256 人分のゲノム情報を解析し、日本人特有の遺伝的特徴を明らかにすることを企図しました。この大規模なデータセットにより、日本人の遺伝的起源、ネアンデルタール人やデニソワ人との遺伝的つながり、そして自然選択のプロセスに新たな光を当てることができました。

研究手法と成果

バイオバンク・ジャパン (BBJ) により、全国 7 地域 (北海道、東北、関東、中部、関西、九州、沖縄) の医療機関に登録された合計 3,256 人分の全ゲノムシーケンス (WGS) を行って、「Japanese Encyclopedia of Whole/Exome Sequencing Library (JEWEL)」というデータセットが作成されました (図 1A)。この最終データセットは、45,586,919 個の SNV^[6]と 9,113,420 個の indel^[7]を含み、そのうち 15,410,953 個 (32.7%) が JEWEL で新たに観察されたバリエーション^[8]です。これは日本における最も包括的な全ゲノムシーケンスデータの一つです。日本人の集団構造を理解するために、まずコモンバリエーション^[8]に基づいて従来の主成分分析 (PCA) ^[9]を行いました。以前の研究と同様に、この分析から沖

縄と本土クラスタの二重構造が再現されました（図 1B）。レアバリエント^[8]が人口構造を明らかにする上でより多くの情報を提供する可能性があるとして仮定し、1,835,116 個のレアバリエントを使用した次元削減（PCA-UMAP）分析^[10]を行いました。この分析により、日本人人口の前例のない遺伝構造が明らかになりました（図 1C）。この構造は、コモンバリエントに基づいて PCA から得られたパターンを再現するだけでなく、いくつかの顕著な特徴を示しています。

具体的には次の通りです。

- 1) 本土の領域間のより明確な分離と、本土クラスタから沖縄クラスタのより明確な区別が観察されます。
- 2) 東北の人々が細長いエリアにクラスタリングされます。

さらに、WGS のデータ ADMIXTURE 分析^[11]による解析結果は、日本人人口は三つの祖先（以下、K1、K2、K3）の混合によって最もよくモデル化できることを示唆しています。K1、K2、K3 はそれぞれ沖縄、東北、関西で最も高まります（図 1D）。K1（沖縄）成分は、南（沖縄に隣接する地域）を除く本土のサブグループで比較的安定した割合の約 12% を維持し、南ではより高い割合の 22% を示しました。K2（東北）と K3（関西）成分は西から東北へ徐々に変化していくことがわかります（図 1D）。

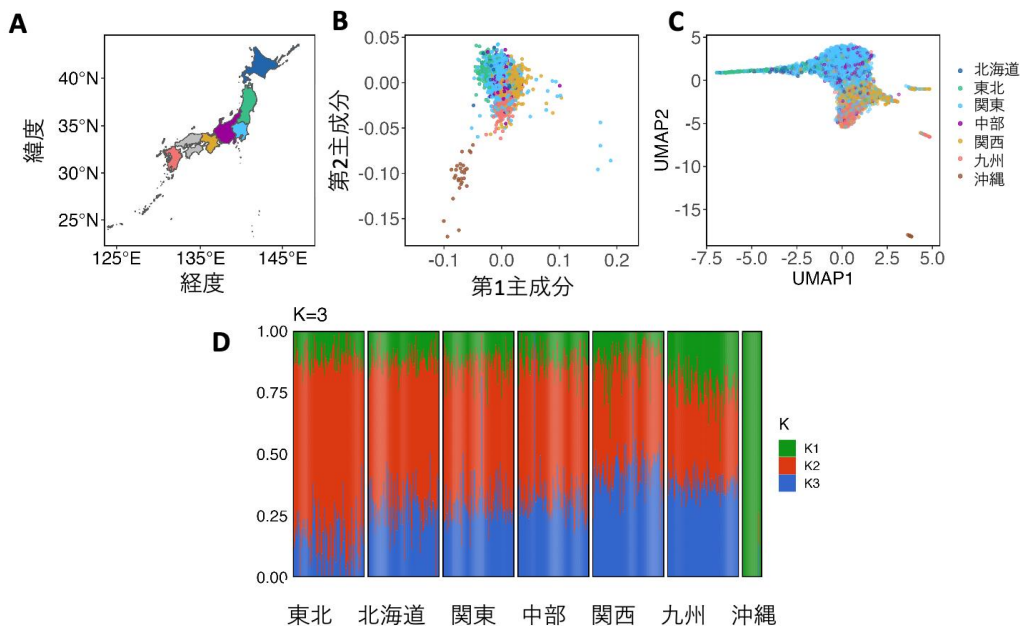


図 1 全ゲノムシーケンスデータを使用した日本人集団構造結果

- A) サンプルが集められた日本の七つの地域。
 - B) コモンバリエントを基にした PCA 分析を行い、参加者を集めた地域ごとに色分けした。
 - C) レアバリエントに基づく PCA-UMAP 分析を示す。
- B)、C) では、一つ一つの点が個人を表しており、点と点との距離は遺伝学的な違いを反映している。近い点ほど遺伝学的には近縁である。この二つの図では、出身地を元にした遺伝背景の違いにより、色分けして表示している。
- D) ADMIXTURE 分析を行い、三つの集団 (K=3) に分けた。沖縄以外の地域からはランダムに 100 人を選

び、沖縄からは全員（28人）をそれぞれ分析した。K1は沖縄、K2とK3はそれぞれ東北と関西で最も高い値を示す。

以前の研究により、日本人が縄文人および東アジア（EA、主に漢民族）の祖先を持つことが示唆されています。最近の古代ゲノムの分析からは、北東アジア（NEA）の祖先の影響も指摘されています。この背景のもと、共同研究グループは、縄文、EA、NEAの現代および古代の遺伝データを今回のデータと共に分析しました。

縄文の祖先比率については、沖縄が最も高い比率（28.5%）を持ち、次いで東北（18.9%）である一方、関西が最も低い（13.4%）と推定されました。これは、縄文人と沖縄の人々の間に高い遺伝的親和性があることを示す以前の研究と一致しています。また、関西地方は漢民族と遺伝的親和性が高いことが明らかになりました。

さらに、共同研究グループは中国、韓国、日本から報告された古代人ゲノムデータを使って、東北と関西の間の遺伝的親和性の違いを評価しました。その結果、関西人と黄河（YR）またはその上流地域の中新石器時代および後新石器時代古代中国集団との間に顕著に密接な関係があることが見受けられました。対照的に、東北地方の個体は、縄文人との遺伝的親和性が顕著に高く、また沖縄の宮古島の古代日本人ゲノム（高い縄文比率を持つ）や韓国三国時代（4～5世紀）の古代韓国人とも高い遺伝的親和性を持つことが示されました。

共同研究グループは、絶滅したネアンデルタール人やデニソワ人から引き継がれた可能性のある遺伝子配列を検出するために、最新の確率的手法IBDmix^[12]を使用しました。JEWELの個々のデータ分析により、ネアンデルタール人由来の約49Mb（メガ・ベース：DNAの長さの単位で1Mbは100万塩基対）とデニソワ人由来の約1.47Mbの遺伝子配列が検出されました。合計で、ネアンデルタール人から引き継がれた可能性が高い3,079セグメント（ゲノム領域）とデニソワ人から引き継がれた可能性の高い210セグメントが特定され、それぞれ772Mbと31.46Mbのゲノムをカバーしています（図2）。

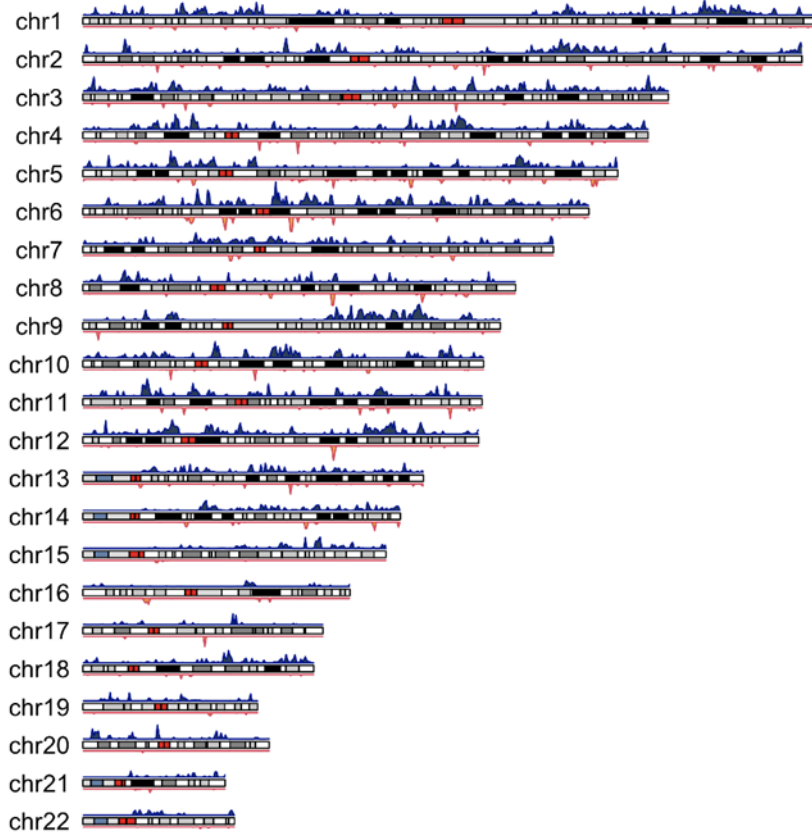


図2 日本人集団におけるネアンデルタール人またはデニソワ人から混入された DNA 配列

各染色体にわたり混入された配列の分布を示す密度プロット。上部のトラック（青で示される）はネアンデルタール人から直接受け継がれたと考えられる配列を表し、下部のトラック（赤で示される）はデニソワ人からの配列を示す。

BBJ から GWAS の結果に基づいて、引き継がれた配列が 106 の表現型に与える影響を調査しました。デニソワ人由来 2 個とネアンデルタール人由来 42 個を含む 44 の領域と 49 の表現型とが関連付けられました。これらのうち 43 の関連は、以前の研究では報告されていません。特に、*POLR3E* のデニソワ人由来セグメントは身長と、*NKX6-1* のセグメントは 2 型糖尿病 (T2D) と関連していました (図 3)。

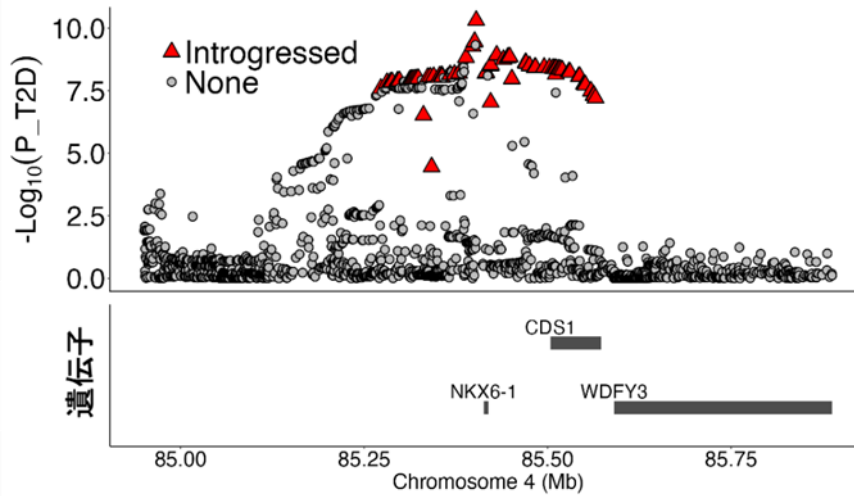


図3 デニソワ人から混入された遺伝子

NKX6-1 遺伝子の領域におけるデニソワ人から直接受け継がれたと考えられる変異は、日本人集団における2型糖尿病 (T2D) と関連している。横軸にヒトゲノム染色体上の位置、縦軸に各変異の T2D との関連の強さを示した。三角形はデニソワ人から受け継がれた変異を指し、灰色の点はそれ以外の変異を示す。

さらに、ネアンデルタール人由来のセグメントは T2D、冠状動脈疾患 (CAD)、安定狭心症 (SAP)、アトピー性皮膚炎 (AD)、グレーブス病 (GD)、前立腺がん (PrCa)、関節リウマチ (RA) など七つの病気と関連する 11 の領域が観察されました。これらの発見は、特に東アジア人において顕著な集団特異性を示しており、日本人におけるアリル頻度^[13]の中央値が欧州人と比較して 21.5 倍であることが分かりました。

共同研究グループは、日本人口において選択された可能性があるゲノム領域を特定するために、全ゲノムスキャンを、iHS^[14]と FastSMC^[15]という二つの方法で実施しました。iHS 法により、共同研究グループは全ゲノム (MHC、ADH クラスター、および ALDH2 を含む) の有意水準で正の選択下にある三つの領域を特定しました (図 4A)。さらに、関西、関東、東北、九州、沖縄の五つの代表的な地域を横断する選択プロファイルの潜在的な地域差を探りました。本州地域全体で類似の選択プロファイルが観察されましたが、アルコール代謝に係る ADH クラスターシグナルは沖縄では比較的弱く、ALDH2 遺伝子領域への自然選択は本土集団でのみ検出されました。これらの差異は、沖縄のサンプルサイズが限られているため、または選択圧が異なる可能性があるため、さらなる研究が必要です。さらに、iHS 法で観察されたシグナルを検証する補完的なアプローチとして FastSMC 法を使用しました (図 4B)。この方法は、過去 50 世代で選択の対象となった可能性がある四つの候補領域を特定しました。これには、iHS 法で有意であった三つの領域 (ADH、ALDH2、MHC) と、候補領域 2p25.3 が含まれます。これら三つの領域 (ADH、ALDH2、MHC) は、以前の研究でも検出されており、日本人口における自己免疫系とアルコール代謝経路への強い選択圧の存在をさらに裏付けるものです。

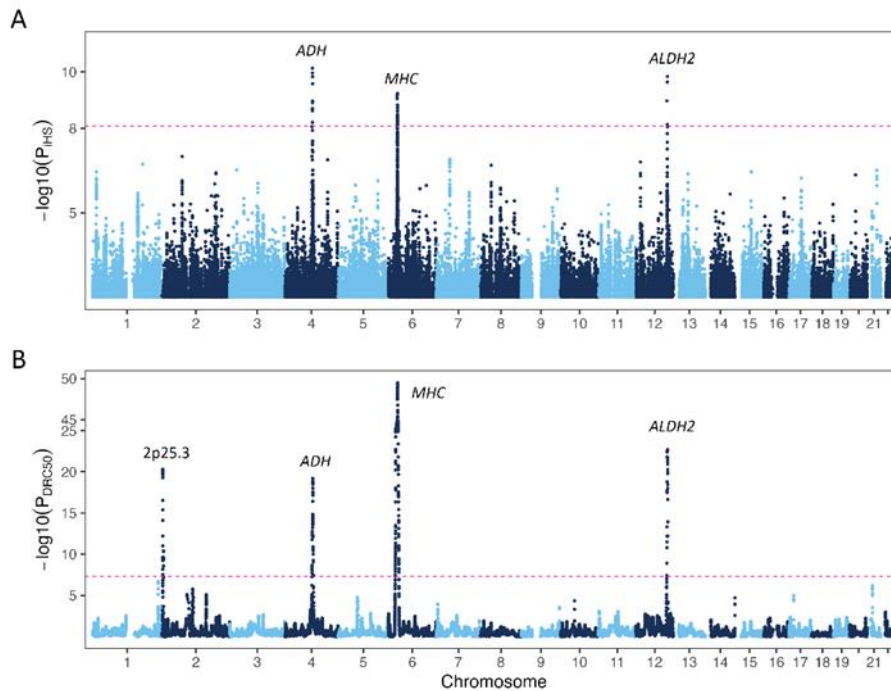


図4 自然選択検出の結果

A: iHS法では、自然選択された可能性のある三つのゲノム領域が特定された。

B: FastSMC法では、自然選択された可能性のある四つのゲノム領域が特定された。

グラフの横軸は染色体上の位置、縦軸は自然選択の痕跡の強さを表す。プロットは各領域の解析結果を示し、縦軸の値が大きいほど強く自然選択の影響を受けたことを意味する。

今後の期待

この研究は、日本人の起源について重要な洞察を提供しています。今まで「二重構造」モデル^[16]、つまり縄文時代の狩猟採集民と大陸からの弥生時代の稲作移民の混血により現代の日本人が形成されたという説は広く受け入れられてきました。最近日本列島の遺跡から出土した人骨のゲノムの研究による「三重構造」モデル^[17]、すなわち、縄文人の祖先集団、北東アジアに起源を持ち弥生時代に日本に渡ってきた集団、そして東アジアに起源を持ち古墳時代に日本に渡ってきた集団の三集団の混血により日本人が形成されたという説が提唱されました。しかし、先行研究で用いられた古人骨全ゲノムのサンプル数は制限されており、より多くの解析が必要と考えられていました。本研究は、大規模な現代日本人ゲノム情報に基づいて、この三重構造モデルの裏付けになり、日本の人口構造をより適切に説明する可能性があると考えられます。さらに、本研究では初めて日本人の遺伝的構造に対する東北地方人の祖先の影響の重要性が強調されました。東北地方は歴史的に蝦夷（エミシ）が居住していた地域であり、彼らの起源を調べる必要があります。

遺伝的分析により、現生人類に非常に近い、絶滅したネアンデルタール人やデニソワ人から引き継がれたDNAが現代日本人の遺伝的多様性にどのように寄与

しているか、またその病気感受性や表現型特性に与える影響はどうか、という点についての理解が深まりました。さらに、*ALDH2* や *ADH* クラスターなど進化的選択を受けた特定遺伝子が日本人の免疫応答やアルコール代謝に重要な役割を果たしていることがすでに示されています。これらの遺伝特性が健康や疾患リスクにどのように影響するかについて、さらなる研究が期待されます。

この研究は、日本における遺伝学研究の新たなマイルストーンを築き、個別化医療の実現に向けた大きな一歩を踏み出すものです。まず希少遺伝変異、特定集団に特有の変異、疾患関連遺伝子の機能解析に貢献します。これらの遺伝的要素を深く理解することで、病気の原因となる遺伝要因の同定が可能になり、新たな治療法の開発ができるようになることが期待されます。さらに、日本人を含む東アジアや世界各地の人々の遺伝的起源と人類史を理解する上で重要な基礎となります。得られた知見は、日本人集団だけでなく、世界の人々に対する個別化医療の発展や新薬の発見にも役立つことが期待されます。将来の研究では、より多くのサンプルと遺伝子解析技術を活用することで、よりよい医療ソリューションの提供につながることを期待されます。

論文情報

<タイトル>

Decoding tri-ancestral origins, archaic introgression, and natural selection in the Japanese population by whole-genome sequencing

<著者名>

Xiaoxi Liu, Satoshi Koyama, Kohei Tomizuka, Sadaaki Takata, Yuki Ishikawa, Shuji Ito, Shunichi Kosugi, Kunihiro Suzuki, Keiko Hikino, Masaru Koido, Yoshinao Koike, Momoko Horikoshi, Takashi Gakuhari, Shiro Ikegawa, Kochi Matsuda, Yukihide Momozawa, Kaoru Ito, Yoichiro Kamatani, Chikashi Terao*

<雑誌>

Science Advances

<DOI>

10.1126/sciadv. adi8419

補足説明

[1] 全ゲノムシーケンス (WGS)

マイクロアレイとは異なり、全ゲノム領域で DNA シーケンスを行い、遺伝子配列を決定する技術。マイクロアレイでは捉えることが難しい頻度の低い SNP の検出などに有用である。次世代シーケンサーの登場と改良により、シーケンスの精度改善とコスト低下が進んでおり、今後さらに需要が高まることが期待される。WGS は Whole Genome Sequence の略。

[2] ネアンデルタール人

ホモサピエンスとの共通祖先から分離した別系統の人類と考えられている。

[3] デニソワ人

ネアンデルタール人の兄弟種である可能性が高いが、現生人類とネアンデルタール人の混血によって生まれたのではないかという議論もある。

[4] 個別化医療

一人一人の体質・病気のタイプに合わせた治療を行うこと。例えば、体質・病気に関連している遺伝子を詳細に調べた上で治療を行う。

[5] バイオバンク・ジャパン (BBJ)

日本人集団 27 万人を対象とした生体試料のバイオバンクで、東京大学医科学研究所内に設置されている。理化学研究所などが取得した約 27 万人のゲノムデータを保有する。オーダーメイド医療の実現プログラム、ゲノム研究バイオバンク事業などを通じて事業が実施され、ゲノム DNA や血清サンプルを臨床情報とともに収集し、研究者へのデータ提供や分譲を行っている。

[6] 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。SNV は single nucleotide variant の略。

[7] 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。

[8] バリエント、コモンバリエント、レアバリエント

バリエントとは、遺伝子の個体ごとの塩基配列の違いのことで、対立遺伝子 (アレル頻度) を保有している割合が高いもの (1%以上) がコモンバリエント、低い (1%未満) ものがレアバリエントである。

[9] 主成分分析 (PCA)

集団遺伝学で用いられる解析手法の一つ。PCA を用いてゲノム全体にわたる大量の SNP (single nucleotide polymorphism: 一塩基多型) の情報を解析すると、各個人が遺伝背景に応じた位置にプロットされるため、個人間の遺伝的な距離を可視化することができる。SNP とは、ヒトゲノム上の個人差の一つで、ヒトゲノムに約 1,000 万カ所存在する一塩基の違い。PCA は principal component analysis の略。

[10] 次元削減 (PCA-UMAP) 分析

データ解析手法の一つ。PCA と組み合わせることで、PCA だけでは区別できない詳細な個人間の遺伝背景の違いを可視化することができる。UPAM は uniform manifold approximation and projection の略。

[11] ADMIXTURE 分析

人々の遺伝的背景が、異なる集団からどのように来ているかを調べる方法。これを使うと、個人の DNA が異なる地理的または民族的集団からどの程度混ざっているかを推定できる。

[12] IBDmix

現代人のゲノム内における古代人の遺伝子の混入を特定するための方法。この方法は、遺伝的に関連する個体間で共有される DNA の断片を探し出すことに基づいている。具体的には、現代人とネアンデルタール人やデニソワ人などの古代人との間で共有される DNA の部分を特定し、現代人の遺伝子に古代人から直接受け継がれた遺伝子がどの程度含まれているかを分析する。IBDmix は identity by descent mixing の略。

[13] アリル頻度

ある集団において特定の SNP が出現する頻度をいう。アリル頻度は人種により大きく異なる。

[14] iHS

自然選択によって有利な遺伝領域が選択されると、その領域の集団頻度が急速に増加する。このようなゲノム領域を検出する方法。iHS は integrated haplotype score の略。

[15] FastSMC

ゲノム全領域における自然選択の痕跡を定量化する解析手法の一つ。各ゲノム領域内に含まれる複数のバリエーションの共通祖先の系譜を高速に探索することで、数十万人規模のゲノム情報に対する自然選択の解析を可能にしている。

[16] 「二重構造」モデル

埴原和郎（ハニハラ・カズロウ）氏による仮説。日本人の祖先集団は、縄文人、そして北東アジアから渡来した弥生人に由来しており、この 2 集団は日本列島内で徐々に混血したが、アイヌ人と南西諸島の人においては、北東アジアからの渡来人の影響は少なかった、とする説。

[17] 「三重構造」モデル

日本人の祖先集団は、縄文祖先、北東アジア祖先、東アジア祖先に由来しており、この 3 集団は日本列島内で徐々に混血したとする説。

共同研究グループ

理化学研究所 生命医科学研究センター

ゲノム解析応用研究チーム

チームリーダー

寺尾知可史（テラオ・チカシ）

（静岡県立総合病院 臨床研究部 免疫研究部長、
静岡県立大学 薬学部ゲノム病態解析講座 特任教授）

上級研究員（研究当時 研究員；静岡県立総合病院 臨床研究部 研究員）

劉 暁溪（リュウ・ギョウケイ）

研究員

石川優樹（イシカワ・ユウキ）

研究員（研究当時）

小杉俊一（コスギ・シュンイチ）

（現 客員研究員）

客員研究員

小池良直（コイケ・ヨシナオ）

上級技師

冨塚耕平（トミヅカ・コウヘイ）

客員研究員	伊藤修司 (イトウ・シュウジ)
ファーマコゲノミクス研究チーム	
特別研究員	曳野圭子 (ヒキノ・ケイコ)
基盤技術開発研究チーム	
チームリーダー	桃沢幸秀 (モモザワ・ユキヒデ)
テクニカルスタッフ (研究当時)	高田定暁 (タカタ・サダアキ)
人材派遣	鈴木邦彦 (スズキ・クニヒコ)
骨関節疾患研究チーム (研究当時)	
チームリーダー (研究当時)	池川志郎 (イケガワ・シロウ)
(現 ゲノム解析応用研究チーム 客員主管研究員)	
循環器ゲノミクス・インフォマティクス研究チーム	
チームリーダー	伊藤 薫 (イトウ・カオル)
特別研究員 (研究当時)	小山智史 (コヤマ・サトシ)
(現 Broad Institute of MIT and Harvard Research Fellow (米国))	
糖尿病・代謝ゲノム疾患研究チーム (研究当時)	
チームリーダー (研究当時)	堀越桃子 (ホリコシ・モモコ)
(現 ゲノム解析応用研究チーム 客員主管研究員)	
東京大学 大学院新領域創成科学研究科	メディカル情報生命専攻
複雑形質ゲノム解析分野	
教授	鎌谷洋一郎 (カマタニ・ヨウイチロウ)
助教	小井土 大 (コイド・マサル)
クリニカルシーケンス分野	
教授	松田浩一 (マツダ・コウイチ)
(東京大学医科学研究所附属ヒトゲノム解析センター	
シーケンス技術開発分野 特任教授、	
バイオバンク・ジャパン (BBJ) 代表)	
金沢大学 人間社会研究域附属 国際文化資源学研究センター	
助教	覚張隆史 (ガクハリ・タカシ)

研究支援

本研究は、日本医療研究開発機構 (AMED) 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) 「先天的/後天的構造多型に着目した免疫/精神疾患病態解明に関する研究開発」「炎症性関節炎の統合ゲノミクス解析」、同難治性疾患実用化研究事業「シングルセル統合ゲノミクス解析が解き明かす強皮症の病態基盤の開発」、同革新的がん医療実用化研究事業「体細胞モザイクのがん発症および予後因子としての意義解明の開発」、同免疫アレルギー疾患実用化研究事業「先天的・後天的ゲノム情報と臨床情報を用いた関節リウマチの層別化基盤の構築」、同ゲノム医療実現バイオバンク利活用プログラム (ゲノム研究バイオバンク) 「利活用を目的とした日本疾患バイオバンクの運営・管理」による助成を受けて行われました。

発表者・機関窓口

理化学研究所 生命医科学研究センター ゲノム解析応用研究チーム

上級研究員 劉 暁溪 (リュウ・ギョウケイ)
(研究当時 研究員; 静岡県立総合病院 臨床研究部 研究員)
チームリーダー 寺尾知可史 (テラオ・チカシ)
(静岡県立総合病院 臨床研究部 免疫研究部長、
静岡県立大学 薬学部ゲノム病態解析講座 特任教授)



寺尾知可史



劉 暁溪

東京大学医科学研究所附属ヒトゲノム解析センター シークエンス技術開発分野
特任教授 松田浩一 (マツダ・コウイチ)

<機関窓口>

理化学研究所 広報室 報道担当

Tel: 050-3495-0247

Email: ex-press [at] ml.riken.jp

静岡県立総合病院 総務課

Tel: 054-247-6111 Fax: 054-247-6140

Email: sougou-soumu [at] shizuoka-pho.jp

静岡県立大学 広報・企画室

Tel: 054-264-5130

Email: koho [at] u-shizuoka-ken.ac.jp

東京大学医科学研究所 プロジェクトコーディネーター室 (広報)

Tel: 090-9832-9760

Email: koho [at] ims.u-tokyo.ac.jp

※上記の[at]は@に置き換えてください。