ヒトゲノムの未解明領域「セントロメア」の大規模解析が実現

世界中に蓄積されたゲノムデータを活用し、がんなどの疾患研究を加速

発表のポイント

- これまで、ヒトゲノムの未解明領域とされ、大規模解析が困難であったヒト染色体「セントロメア」について、既存のショートリード全ゲノム解析（WGS）データを活用して解析できる新しい手法を開発しました。
- 開発した新手法で、世界 3,332 人のショートリード WGS データを解析し、セントロメアにも人類集団ごとの多様なタイプが存在することを明らかにしました。
- さらに、脳腫瘍（乏突起膠腫）を解析した結果、セントロメアの個人差が、乏突起膠腫に特徴的な染色体異常の形成に関係する可能性を示しました。
- 本研究により、世界中に蓄積されたショートリード WGS データを活用してセントロメアを大規模に調べることが可能となり、がんをはじめとするさまざまな疾患とセントロメアとの関係の解明が期待されます。

概要

国立研究開発法人国立がん研究センター（理事長：間野 博行、東京都中央区）研究所 ゲノム解析基盤開発分野長の白石友一らの研究グループは、京都大学、名古屋大学をはじめとする国内外の研究機関との共同研究により、これまで、ヒトゲノムの未解明領域とされ、大規模な解析が困難であったヒト染色体の「セントロメア[注1]」について、ショートリード全ゲノム解析（WGS）[注2] データから個人ごとのタイプを推定する手法を開発しました。

セントロメアは、染色体の中央部分に位置し、細胞分裂の際に紡錘糸が結合して染色体を二つの細胞へ正確に分配するために不可欠な領域です。しかし、セントロメアは同じようなDNA配列が長く繰り返しているため、一般的なゲノム解析では詳しく調べることが難しく、これまで十分な解析が行われてきませんでした。近年、長い DNA 配列を読み取るロングリードシークエンス[注3] によって詳細な解析が可能になってきましたが、多くの DNA 量を必要とし、シークエンスコストも高いため、数千人、数万人規模の研究に用いることは困難でした。

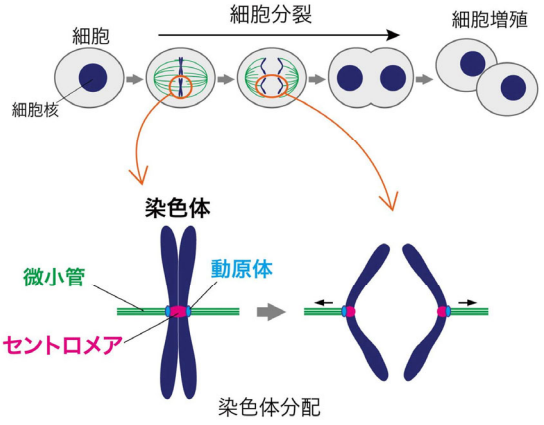

図1：セントロメアの位置と役割

出典：大阪大学 深川竜郎研究室ウェブサイト（一部改変）

研究グループは、よく似た配列が繰り返すセントロメアの中にも、タイプによって異なる短い DNA 配列があることに着目し、この短い DNA 配列を目印として、セントロメアを配列や構造が似たタイプに分類しました。この原理を利用し、ショートリード WGS データからセントロメアのタイプを推定するソフトウェア「ascairn（エーエスケルン）」を開発しました。

開発した新手法によって、世界各地の 3,332 人のゲノムデータを解析した結果、セントロメアには多様

なタイプがあり、個人差のほかに地域や集団による違いがあることが分かりました。さらに、脳腫瘍の一種である乏突起膠腫[注4]を解析した結果、乏突起膠腫に特徴的な「1p/19q 共欠失」という染色体異常の形成に、セントロメアの個人差が関連することを示しました。

本研究により、世界中ですでに蓄積されているショートリード WGS データを活用して、セントロメアを大規模に調べることが可能になります。今後、人類のゲノムの多様性に加え、がんをはじめとする疾患で染色体異常が生じる仕組みの解明につながることが期待されます。

本研究は、国立がん研究センター研究所 ゲノム解析基盤開発分野長 白石友一、京都大学大学院医学研究科 腫瘍生物学講座 教授 小川誠司（兼：同高等研究院ヒト生物学高等研究拠点（WPI-ASHBi）主任研究者）、助教 越智陽太郎、京都大学高等研究院 ヒト生物学高等研究拠点（WPI-ASHBi）特定准教授 辻村太郎、京都大学 iPS 細胞研究所 未来生命科学開拓部門 教授 山本拓也（兼：同高等研究院ヒト生物学高等研究拠点（WPI-ASHBi）主任研究者）、名古屋大学大学院医学系研究科 脳神経外科学 教授 齋藤竜太らによる、国内外の研究機関との共同研究です。

本研究成果は、米国科学誌「*Cell*」に米国東部時間 2026 年 10 月 13 日（日本時間 2026 年 10 月 14 日）に掲載されました。

背景

ヒトの細胞には通常 46 本の染色体があり、それぞれの染色体には「セントロメア」と呼ばれる領域があり、細胞分裂の際に、複製された染色体を二つの細胞に分配する重要な役割を担っています。

ヒトのセントロメアは、同じような DNA 配列が長く繰り返す構造をしています。一般的なショートリードシークエンスでは DNA を短い断片に分けて読み取るため、読み取った断片が反復配列のどこに由来するかを判別することが困難です。このため、セントロメアは、ヒトゲノムの中でも特に解析が難しい領域として残されてきました。

近年、長い DNA 配列を読み取るロングリードシークエンス技術の発展と、ヒトのゲノム多様性を包括的に明らかにするヒトパンゲノムプロジェクト[注5]などの進展により、セントロメアの詳しい配列や、個人差も徐々に明らかになってきました。しかし、セントロメアの大規模解析には、依然として高いコストと高度な解析技術が必要です。

一方、世界中では、DNA を短い断片として読み取ったショートリード WGS データが大量に蓄積されています。そこで研究グループは、これらの既存データを活用してセントロメアを調べることができれば、人類集団における多様性や、がんなどの疾患でみられる染色体異常との関係を、数千人から数十万人という規模で研究できると考えました。

研究成果（方法・結果）

<u>短い DNA 配列を「目印」にセントロメアのタイプを分類</u>

研究グループは、118 人分の高品質なゲノム配列からセントロメアの配列を集めました。次に、繰り返し配列の中にある 27 塩基の短い DNA 配列に着目しました。本研究では、このうちセントロメアのタイプを見分ける目印となるものを「rare k-mer[注6]」と定めました。

短い DNA 配列の有無や組み合わせを比べると、セントロメアの構造や由来の違いを見分けられることが分かりました。この目印を使い、各染色体のセントロメアをそれぞれ 3～16 のタイプに分類しました。全染色体では計 214 タイプに分かれました。

既存の全ゲノム解析データからセントロメアを調べる新手法を開発

研究グループは、全ゲノム解析データに含まれる短い DNA 配列の特徴を利用して、セントロメアを解析するソフトウェア「ascairn（エーエスケルン）」を開発しました。この手法は、すでに蓄積されているショートリード WGS データを利用して、セントロメアのタイプを推定します。

性能評価では、染色体ごとのタイプの組み合わせを80.7～100%の精度で推定でき、精度の中央値は92.7%でした。なお、現段階では臨床診断に用いる検査ではなく、集団規模のゲノム研究や疾患研究に用いる解析手法です。

これまでセントロメアを詳しく調べるには特別な解析が必要でしたが、本手法によって世界中に蓄積されている大量のショートリード WGS データを活用した大規模研究が可能になります。

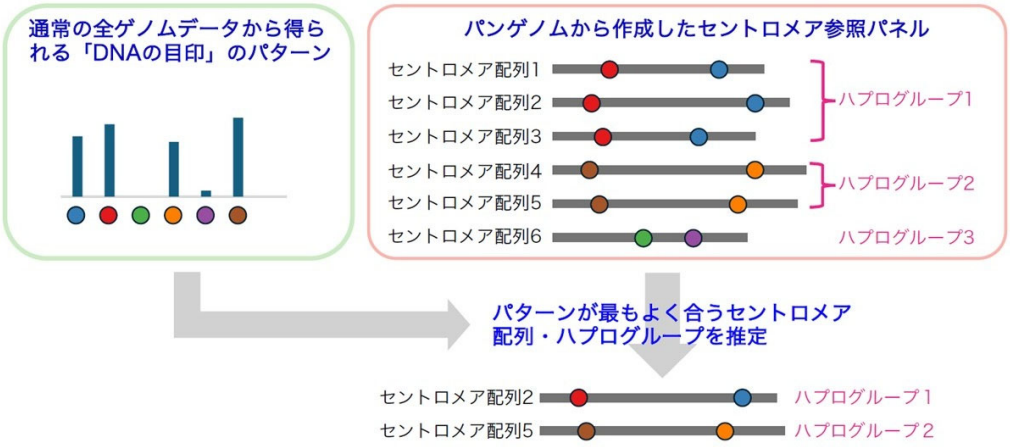

図 2　ショートリード WGS データからセントロメアのタイプを推定

ヒトパンゲノム研究などから得られた高品質なセントロメア配列をもとに参照パネル[注7]を作成し、ショートリード WGS データ中の rare k-mer のパターンから、各染色体のセントロメア・ハプログループ[注8]を推定します。

世界 3,300 人超の解析からセントロメアの多様性と地域差を確認

世界各地の計 3,332 人のショートリード WGS データ（1000 Genomes Project の 2,504 人と Human Genome Diversity Project の 828 人）を解析したところ、セントロメアのタイプの分布には地域差がありました。特にアフリカに由来する解析対象集団では、他の解析対象集団ではほとんどみられないタイプが確認されました。こうした分布は、現生人類の分岐や移動の歴史と整合していました。ゲノムの他の領域と同じようにセントロメアも、ヒト集団の分岐や移動などの歴史を反映しながら多様化してきたと考えられます。

本研究でセントロメアの大規模解析が可能となったことで、今後は、生まれつきのセントロメアの違いが、がんをはじめとする疾患とどのように関係するのかを大規模に調べることが可能になります。

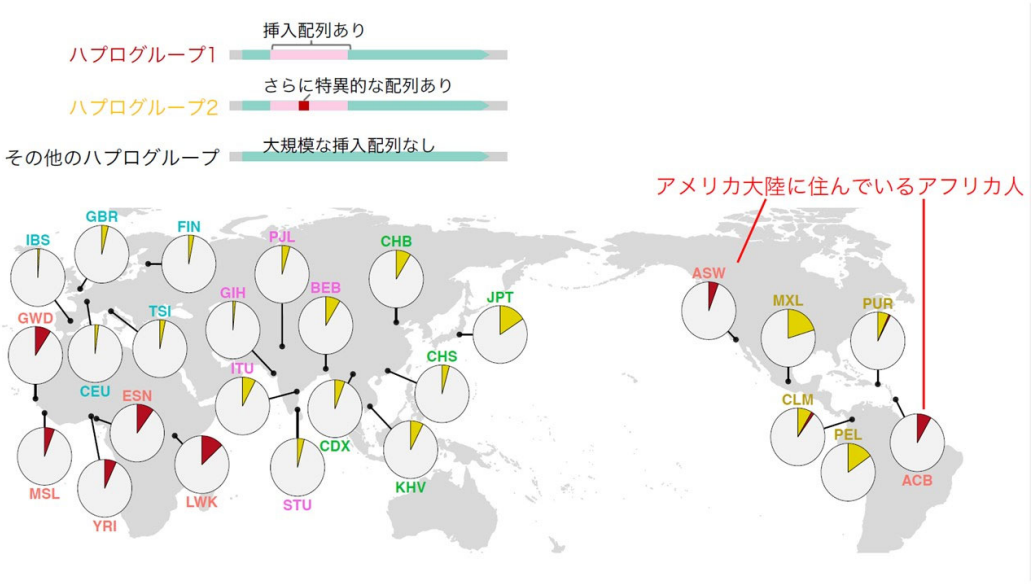

図 3　世界各地で異なるセントロメア・ハプログループの分布

世界各地の集団でセントロメア・ハプログループの頻度を比較したところ、明瞭な地域差が認められました。図は 19 番染色体のセントロメアの例を示しています。

<u>がんに特徴的な染色体異常の形成にセントロメアが関与する可能性を発見</u>

研究グループは、開発した手法をがんのゲノム解析に応用しました。がんでは、染色体の一部が失われたり、別の染色体とつながったりする「染色体異常」が起こることがあります。しかし、その異常がどこで、どのように生じるのかは十分に分かっていない場合があります。

そこで、脳腫瘍の一種である乏突起膠腫でみられる「1p/19q 共欠失」と呼ばれる特徴的な染色体異常について、がんゲノムデータベース、TCGA[注 9]に登録された 142 例のショートリード WGS データを用いて解析しました。その結果、1 番染色体と 19 番染色体の切断点がセントロメア内部に位置し、そこで両染色体が転座[注 10]することを明らかにしました。また、切断位置は、動原体形成領域を境に、1 番染色体と 19 番染色体でそれぞれ特定の側に偏っていることが分かりました。さらに、名古屋大学大学院医学系研究科脳神経外科学から提供された 2 症例をロングリードシークエンスで解析し、この転座構造を直接確認しました。

さらに、この 142 例の解析の中で、同じ人が持つ同じ番号の 2 本の染色体のうち、どちらのセントロメアが実際の転座に関与したかを調べました。その結果、動原体形成領域からセントロメア端までの「転座が起こり得る範囲」が長いハプログループほど、実際の転座に関与しやすいことが分かりました。

これらの結果は、これまで解析が難しかったセントロメアが、がんに特徴的な染色体異常の形成に関与していることを示すもので、今後は、さまざまながんでみられる染色体異常がどのように生じるのかを理解する手掛かりになることが期待されます。

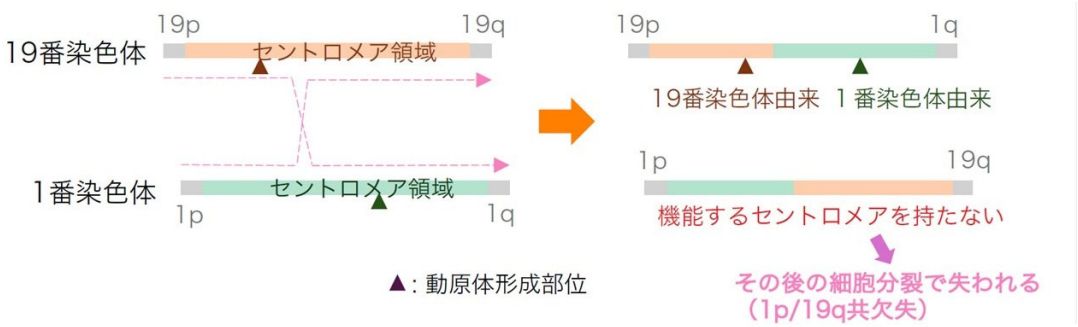

図 4　セントロメアの構造的な個人差と 1p/19q 共欠失の形成

1 番染色体と 19 番染色体はセントロメア内部で転座し、その後、一方の染色体が失われることで 1p/19q 共欠失が形成されます。動原体形成領域[注 11] からセントロメア端までの「転座が起こり得る範囲」には個人差があり、この範囲が長いハプログループほど転座に関与しやすいことが示されました。

展望

これまでの大規模ゲノム研究では、遺伝子や比較的解析しやすいゲノム領域が主な研究対象となり、セントロメアをはじめとする高度反復領域の情報は十分に活用されてきませんでした。今回開発した手法によって、世界中ですでに蓄積されているショートリード WGS データから、これまで見えにくかったセントロメアの情報を新たに取り出すことが可能になります。

今後、より多様な集団を含む数万人から数十万人規模のゲノムデータへ解析を広げることで、セントロメアの多様性と人類の進化・集団形成との関係を、さらに詳しく調べることができます。また、他のがんや染色体異常にも対象を広げ、セントロメアの構造的な個人差が疾患にどのように関与するかを検証することが期待されます。ヒトゲノムの「読みにくい領域」にまで解析対象を広げることで、従来のゲノム解析では捉えられなかった染色体異常の形成機構の理解につながることが期待されます。

論文情報

雑誌名：　*Cell*

タイトル：　Rare k-mers reveal centromere haplogroups underlying human diversity and cancer translocations

著者：　Yuichi Shiraishi*, Yotaro Ochi, Masahiro Sugawa, Yoshitaka Sakamoto, Keisuke Kimura, Taro Tsujimura, Ai Okada, Rurika Okuda, Shinichi Namba, Tetsuya Hori, Yoshihisa Kobayashi, Tsubasa Miyauchi, Raul N. Mateos, Hajime Suzuki, Kenichi Chiba, Yu Ito, Wataru Nakamura, Fumiharu Ohka, Kazuya Motomura, Takuya Yamamoto, Yosuke Kawai, Yukinori Okada, Hiromichi Suzuki, Tatsuo Fukagawa, Motohiro Kato, Ryuta Saito, Erik Garrison, Glennis A. Logsdon, Seishi Ogawa (*Corresponding author)

DOI：　10.1016/j.cell.2026.08.053

掲載日：　2026 年 10 月 13 日

URL：　https://doi.org/10.1016/j.cell.2026.08.053

研究費

- 国立研究開発法人日本医療研究開発機構（AMED）
  研究事業名：次世代がん医療加速化研究事業
  研究課題名：パンゲノム参照配列を用いた難読領域におけるがんドライバー変異の網羅的探索
  研究代表者名：白石　友一（国立がん研究センター研究所）

  研究事業名：生命科学・創薬研究支援基盤事業（BINDS）
  研究課題名：ロングリード 1 分子エピゲノム解析の支援
  研究代表者名：辻村　太郎（京都大学）

- 国立がん研究センター
  研究事業名：国立がん研究センター研究開発費
  研究課題名：日本人完全版ゲノムリファレンス配列の構築とがんゲノム解析への応用
  研究代表者名：白石　友一（国立がん研究センター研究所）

用語解説

注 1　セントロメア

染色体に存在し、細胞分裂の際に、複製された染色体を二つの娘細胞へ正確に分配するために不可欠な領域です。細胞分裂時には、セントロメアの一部に「動原体」と呼ばれる構造が形成され、染色体を移動させるための足場となります。

注 2　全ゲノム解析（WGS）

ヒトが持つ DNA 全体（ゲノム）の配列を網羅的に調べる解析です。本研究では、DNA を短い断片に分けて読み取る一般的なショートリード WGS データを用いました。

注 3　ロングリードシークエンス

一度に長い DNA 配列を読み取る技術です。繰り返し配列が多いセントロメアなど、ショートリードシークエンスでは解析が難しい領域の詳細な配列や構造を調べることに適しています。

注 4　乏突起膠腫

乏突起膠腫は、脳に発生する神経膠腫の一種です。染色体はセントロメアを境に、短い側の「短腕」と長い側の「長腕」に分かれ、短腕は p、長腕は q で表します。1 番染色体の短腕（1p）と 19 番染色体の長腕（19q）が同時に失われる「1p/19q 共欠失」は、乏突起膠腫の診断上重要な特徴です。

注 5　ヒトパンゲノム／ヒトパンゲノムプロジェクト

一人のゲノム配列だけでは捉えきれない人類の遺伝的な多様性を明らかにするため、多様な由来を持つ複数の人のゲノム配列を集めて構築したゲノム情報です。ヒトパンゲノムプロジェクトでは、従来の標準的なゲノム配列では十分に表現できなかった個人差や複雑なゲノム領域の解析が進められています。

注 6　rare k-mer

DNA 配列を一定の長さで切り出した短い配列を k-mer と呼びます。本研究では、27 塩基の k-mer を使用しました。その中でも、一部のセントロメアに特徴的に存在し、セントロメアのタイプを識別するために有用な k-mer を rare k-mer として、「DNA の目印」に利用しました。

注 7　参照パネル

解析の比較対象として用いる、複数の人の高品質なゲノム配列をまとめたデータです。本研究では、ロングリードシークエンスなどによって詳しく読み取られたセントロメア配列を参照パネルとして使用し、解析対象者のセントロメアがどのタイプに近いかを推定しました。

注 8　ハプログループ

ハプログループとは、配列や構造が似ており、共通の由来を持つと考えられるハプロタイプをまとめたグループです。本研究では、rare k-mer の組み合わせが似たセントロメアをまとめ、各染色体のセントロメアを複数のハプログループに分類しました。

注 9　TCGA（The Cancer Genome Atlas）

米国で実施された大規模ながんゲノム解析プロジェクト。さまざまながんのゲノムデータを収集し、研究者に提供している。

注 10　染色体転座

染色体転座とは、染色体の一部が切れ、別の染色体などにつながる変化です。染色体が切れ、転座の境界となった位置を「切断点」と呼びます。本研究では、1 番染色体と 19 番染色体の切断点がそれぞれのセントロメア内部にあり、両染色体がそこでつながることを示しました。この転座によって生じた染色体の一方が失われることで、1p/19q 共欠失が形成されると考えられています。

注 11　動原体形成領域

動原体は、細胞分裂の際に染色体を二つの娘細胞へ分配するため、セントロメア上に形成される構造です。本資料では、セントロメアのうち動原体が形成される部分を「動原体形成領域」としています。動原体形成領域を持たない染色体は、細胞分裂の際に正しく分配されず、失われる可能性があります。