

取 扱 注 意		
解 禁	テレビ・ラジオ・ 通信社・インターネット	日本時間 7 月 31 日（金） 午前 9 時以降
	新聞	日本時間 7 月 31 日（金）夕刊

令和 2 年 7 月 30 日
横 浜 市 立 大 学
日本医療研究開発機構

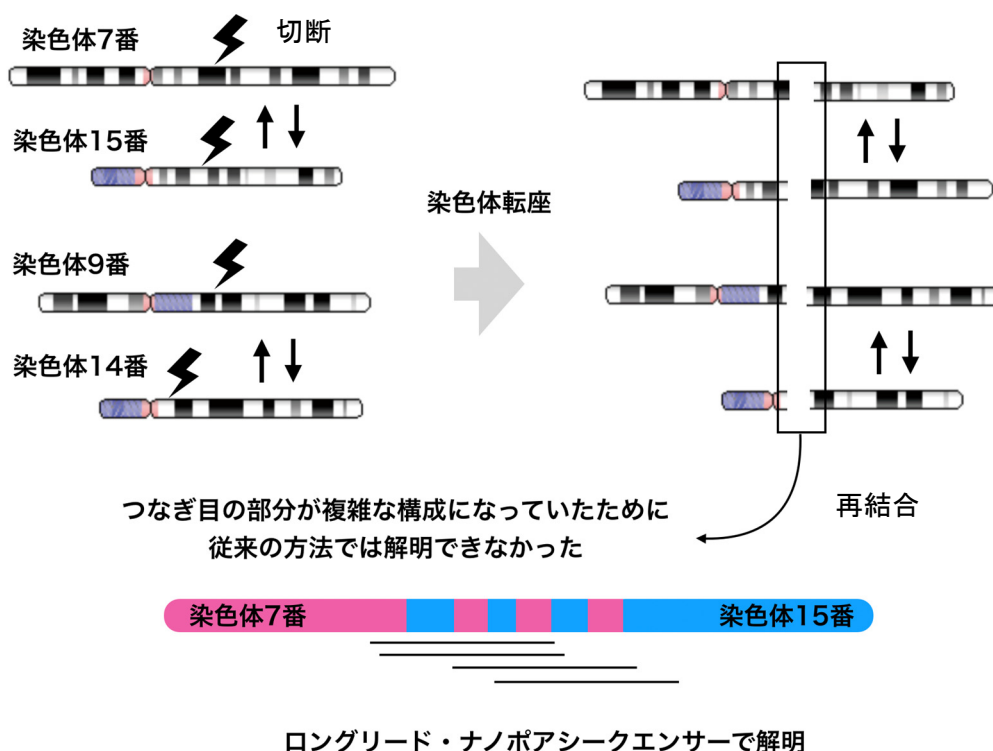
新技術ロングリード・シーケンサーで 複雑なゲノムの構造変化を解明する手法を開発

横浜市立大学大学院医学研究科 遺伝学 三橋里美助教（現 東京医科歯科大学難治疾患研究所 准教授）、尾堀佐知子医師、松本直通教授らの研究グループは、ロングリード・シーケンサー^{*1}を用いた新しいデータ解析手法を開発、これを応用し複数の症例で疾患に関わる複雑なゲノム構造異常を完全に解き明かしました。この研究で用いた新規解析手法は、国立研究開発法人産業技術総合研究所 人工知能研究センター マーティン・フリス上級主任研究員、大阪大学微生物病研究所 遺伝情報実験センター ゲノム情報解析分野 加藤和貴准教授との共同研究で開発されました。

本研究は、『Genome Medicine』に掲載されます。（日本時間 7 月 31 日午前 9 時付オンライン）

研究成果のポイント

- 新技術ロングリード・シーケンサーによる全ゲノム解析から、複雑なゲノム構造異常を迅速かつ効率的に検出する手法を開発
- 非常に複雑に絡み合ったゲノム構造の変化である染色体破砕^{*2}（クロモスリプシス）の全体構造を明らかにした
- 新手法による複雑なゲノム構造変化の解明が、遺伝性疾患の病態解明に繋がることが期待される



（図 1）染色体転座^{*3}に伴う複雑なゲノム構造変化を、ロングリード・ナノポアシーケンサー^{*4}を用いた新しいデータ解析手法で解明。（染色体図: <https://www.ncbi.nlm.nih.gov/genome/tools/gdp>）

研究の背景

遺伝性疾患の原因となるゲノム変化の解明には、DNA 塩基配列⁵を直接読む方法（DNA シークエンサー）や、マイクロアレイなどが汎用されてきました。遺伝性疾患の原因となるゲノム異常は、DNA の塩基変化（SNV: single nucleotide variant）と大きな領域のコピー数が変わる変化（CNV: copy number variant）が知られていますが、そのほかにも、染色体転座³や、染色体の構造が複雑に変化する染色体破砕²（クロモスリプシス、chromothripsis）と呼ばれる現象など、様々なものがあり、構造多型⁶と呼ばれています。

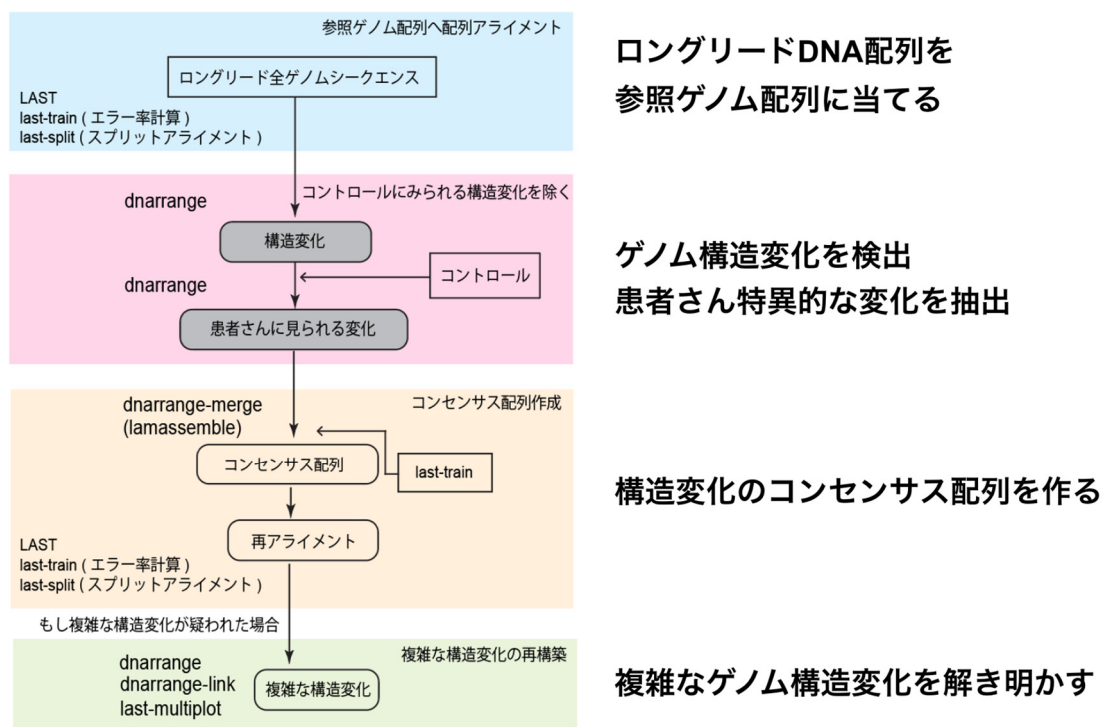
従来、網羅的に DNA 塩基配列を解析するにはショートリード・シーケンサーと呼ばれる 200 塩基程度の配列を一度に読むことのできる装置を使うことが一般的ですが、近年この 100 倍もの長さを読むことのできるロングリード・シーケンサーが実用化され、疾患のゲノム解析にも応用されるようになってきています。

長い配列を読むことができるようになると、染色体の広範な領域の構造変化をより正確に検出できるようになると期待されています。しかし、ロングリードのデータから、複数の個人での構造変化を比べることや、複雑な構造変化を検出することはまだ効率的な手法が限られ容易ではなく、新しい手法の登場が待ち望まれていました。

研究の内容

従来の方法では解明が困難であった染色体転座を持つ先天奇形や卵巣機能不全等の症状のある患者さん 4 例について、ロングリード・ナノポアシーケンサーを用いて全ゲノム解析を行いました。本邦で開発された配列アライメント⁷のデータ解析ツールである LAST (<http://last.cbrc.jp>) を用いて、参照ヒトゲノム配列にアライメントを行い、共同研究により新規に開発した dnarrange（ディーエヌアレンジ）というソフトウェアを用いることで、構造変化を検出しました（図 2）。

データ解析パイプライン



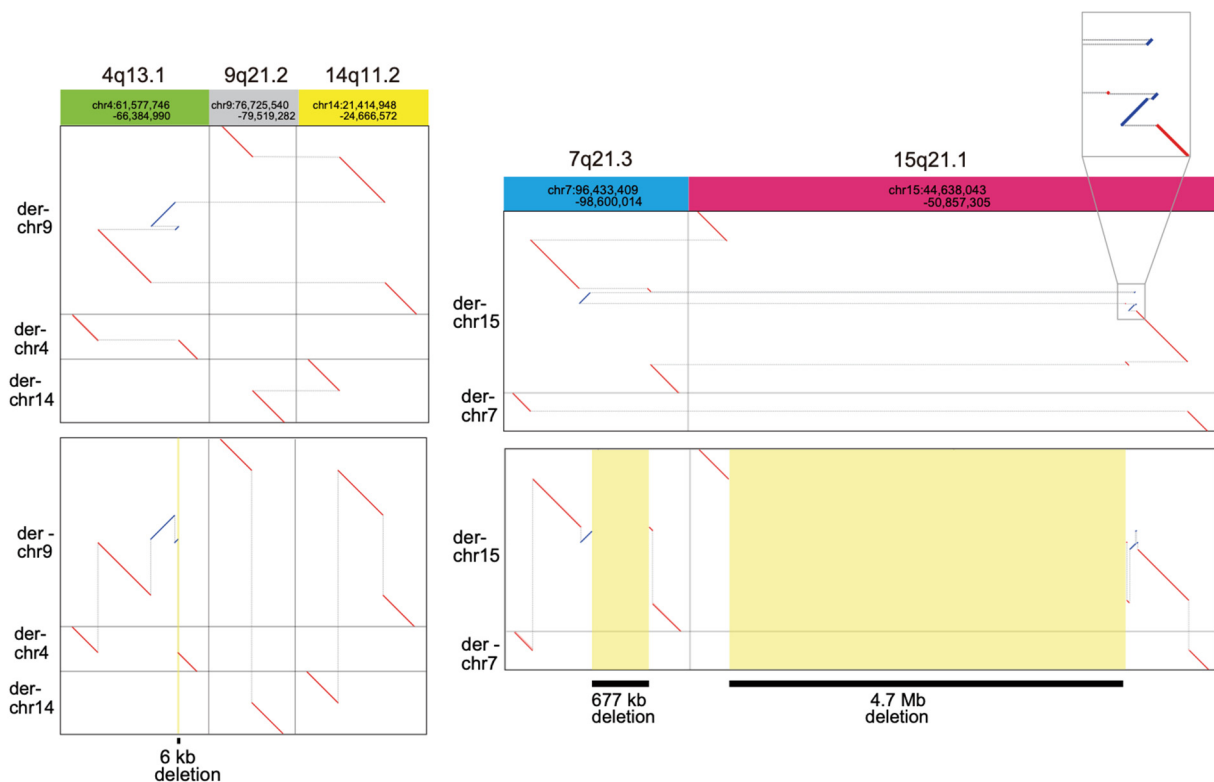
（図 2）データ解析パイプライン⁸の全容。ロングリード・シーケンサーのデータから、複数の解析手法を組み合わせ、患者さんのゲノム構造変化を明らかにする。まず、ロングリードの DNA 塩基配列を参照ゲノム配列と比較（図：ブルー部分）。ゲノム構造変化を抽出、さらに他のゲノム（コントロール）が持たず、患者さんに特異的な変化を抽出する（図：ピンク部分）。構造変化のコンセンサス配列を作り、切断点を詳細に解析する（図：オレンジ部分）。さらに、複雑な構造変化を再構築する（図：グリーン部分）。

ヒトのゲノムには疾患を引き起こさない構造変化が数千個以上存在しています。これらをうまく除外することにより、疾患と関係する構造変化をより効率よく見つけることができると考えられます。今回、同じ疾患を持たない個人のゲノム（コントロール）に存在する構造多型を、**dnarrange** を用いて除外し、患者さんだけで見られる病的意義を持つ候補となる構造異常領域を 100 個以下に絞ることができました。

1 例の患者さんでは、5 本の染色体が 18 個の切断点を持ち、非常に複雑な構造変化をきたしていました。この患者さんのゲノムが実際にどのように変化しているのかを知ることは、従来法では極めて困難でした。そこで本研究では、これらの構造多型の位置関係を、自動で推測するツール **dnarrange-link** を開発することにより、この複雑なゲノム構造異常を再構成することができました。この再構成によって、染色体破砕により 3 箇所の広い領域が欠失していることがわかりました（図 3）。これらの欠失は、ゲノム構造を完全に再構成することで初めて描出することが可能となるため、従来の切断点を検出する手法だけで欠失したゲノム領域を同定することは困難であったと考えられます。

また、構造変化の影響やメカニズムを調べるには、ゲノム上の切断点を知ることが重要です。しかし、現在のロングリード・シーケンサーから出されるデータはエラーが多いため、一塩基の精度で正確な配列を知ることが困難な場合があります。そこで多重アライメントソフトウェア **MAFFT** (<https://mafft.cbrc.jp/alignment/software/>) を用いて、コンセンサス配列^{*9}を作成する手法 **lamassemble**（ラマセンプル）を開発し、構造変化部位のコンセンサス配列を得ました（図 2）。コンセンサス配列から明らかになった構造多型の切断点は、サンガーシーケンス^{*10}により確認された切断点と 0~1 塩基しかずれておらず、極めて高精度であり、染色体破砕をきたしたゲノム構造異常のメカニズムを考察することが可能となりました。

5つの染色体が関係する非常に複雑なゲノム構造変化を検出



（図 3）1 例では複数の構造多型が非常に複雑に絡み合っていたが、本研究で全体像を見ることができた（上図）。全体像を構築することによって初めて明らかとなった欠失部位は黄色で示す（下図、5 本の染色体に 18 個の切断点が存在する。切断点は灰色の横線で繋がった部分で示した。赤線は参照ゲノム配列に対して順行であり、青線は逆転がおきていることを示す）。

今後の展開

本研究により、従来の手法で解決困難であった染色体転座の患者さんのゲノム配列を、新技術であるロングリード・シーケンサーで読んだデータを用いて、疾患と関係するゲノム構造変化を明らかにする手法を確立しました。従来の手法では解析が難しい、染色体破碎のような非常に複雑な構造変化が解き明かされるようになり、遺伝性疾患の病態解明に繋がることが期待されます。

用語説明

- *1 ロングリード・シーケンサー: 1 万塩基以上の DNA 配列を一つづきに読むことができる装置。Oxford Nanopore Technologies や Pacific Biosciences 社の装置が使われる。
- *2 染色体破碎 (クロモスリプシス): 染色体が複数の断片に破碎され、再結合されることでゲノムに複雑な構造変化が起きること。遺伝子領域が欠損したり、遺伝子の発現に影響を与えたりすることで、がんや先天性疾患を引き起こすことが知られている。
- *3 染色体転座: 染色体の一部が切断され、異なる染色体と再結合して入れ代わる変化。
- *4 ナノポアシーケンサー: Oxford Nanopore Technologies 社のロングリード・シーケンサー。ナノポアというタンパク質の穴が人工膜に埋め込まれており、ナノポアを DNA 分子が通り抜ける時に起きる電流の変化によって、DNA の配列を解析することができる。
- *5 DNA 塩基配列: ゲノムを構成する DNA 塩基配列は、グアニン (G)、シトシン (C)、アデニン (A)、チミン (T) の 4 種類からなる。
- *6 構造多型: ゲノムの比較的大きな領域に起きる変化。欠失・重複・逆位・転座・挿入など、複数のパターンに分類される。量の変化を伴うものは、CNV (copy number variation) と呼ばれる。
- *7 配列アライメント: 複数の DNA などの配列を、似たものが同じ位置に来るように並べたもの、またその手法。
- *8 解析パイプライン: データ解析は複数の手法を使って行われることが多いが、これらを一連の流れとして組み合わせたもの。
- *9 コンセンサス配列: 配列アライメントの各位置において最も確からしい塩基を計算し、1 種類の配列を決めること。
- *10 サンガーシーケンス: 1970 年代にフレデリック・サンガーが開発した塩基配列の決定法で、DNA ポリメラーゼを用いて、末端が特定の塩基に対応する DNA 断片を合成し、一つひとつ DNA 配列を決めていく。

掲載論文

A pipeline for complete characterization of complex germline rearrangements from long DNA reads

Satomi Mitsuhashi, Sachiko Ohori, Kazutaka Katoh, Martin C Frith, Naomichi Matsumoto

Genome Medicine (2020) DOI: 10.1186/s13073-020-00762-1

※本研究は、国立研究開発法人日本医療研究開発機構 (AMED) の難治性疾患実用化研究事業「新技術を用いた難治性疾患の高精度診断法の開発」(研究代表者: 松本直通)、厚生労働省、日本学術振興会、武田科学振興財団の支援により実施されました。

※本研究の共同研究により開発されたツールは以下になります。

dnarrange: <https://github.com/mcfrith/dnarrange>

lamassemble: <https://gitlab.com/mcfrith/lamassemble>

＜お問い合わせ先＞



（本資料の内容に関するお問い合わせ）

公立大学法人横浜市立大学大学院医学研究科 遺伝学 教授 松本直通

TEL : 045-787-2606 FAX : 045-786-5219

E-mail : naomat@yokohama-cu.ac.jp

（取材対応窓口、詳細の資料請求など）

公立大学法人横浜市立大学 研究・産学連携推進課長 山崎理絵

TEL : 045-787-2510

E-mail : kenkyupr@yokohama-cu.ac.jp



（AMED の事業について）

国立研究開発法人日本医療研究開発機構

ゲノム・データ基盤事業部医療技術研究開発課

TEL : 03-6870-2221

E-mail : nambyo-r@amed.go.jp

URL : <http://www.amed.go.jp/>