

長鎖シーケンサーを用いた 日本人の遺伝的多様性と癌の体細胞変異の包括的解析

1. 発表者：

藤本 明洋（東京大学大学院医学系研究科国際保健学専攻 教授）

2. 発表のポイント：

- ◆長鎖シーケンサー（注1）のデータからヒトの構造異常（遺伝的多様性と体細胞変異）を検出する手法を開発した。
- ◆11人の日本人サンプル由来の正常組織および癌の全ゲノムシーケンス（注2）を行った結果、遺伝的多様性と体細胞変異が生じるメカニズムの違いが示唆された。
- ◆本研究で開発した手法は、遺伝性疾患の原因遺伝子の探索やがんの変異解析にも利用可能であり、貢献が期待できる。

3. 発表概要：

近年の次世代シーケンサー（注3）に代表される塩基配列決定技術の発展により、がんや遺伝性疾患の原因変異の探索が行われている。しかし、逆位や転座などの構造異常（注4）の検出は困難であり、ヒトの構造異常の全体像は未だ解明されていない。東京大学大学院医学系研究科の藤本明洋教授らは、長鎖シーケンサー（Oxford Nanopore）のデータから構造異常を検出する手法（CAMPHOR ソフトウェア）を開発した。それを用いて、日本人の遺伝的多様性（遺伝的な個人差）とがんの体細胞変異を対象とし構造異常を検出したところ、遺伝的多様性における挿入・欠失の特徴に関しては挿入の9割はトランスポゾン（注5）由来である一方、欠失にはそのような特徴がないことが明らかになった。さらに構造異常の切断点の詳細な解析から、構造異常の原因を推定した結果、短鎖次世代シーケンサーを用いた先行研究（引用文献1）では検出されていない非相同組み替え（注6）により生じた構造異常が発見された。

本研究で開発した手法は、遺伝性疾患の原因遺伝子探索やがんの変異解析に貢献すると考えられる。

4. 発表内容：

近年の次世代シーケンサーに代表される塩基配列決定技術の発展により、がんや遺伝性疾患の原因変異探索が行われている。しかし、次世代シーケンサーは決定できる塩基の長さ（リード長）が短く、逆位や転座などの構造異常の検出は困難であると考えられていた（図1）。近年、リード長が長い長鎖シーケンサーが開発され、遺伝学研究への導入が進みつつある。長鎖シーケンサーを用いることで、次世代シーケンサーでは検出が困難な多型や変異が検出されると期待される。しかし、長鎖シーケンサーはエラー率の高さなどの問題のため、多型や変異の検出の手法は確立していない。東京大学大学院医学系研究科の藤本明洋教授らは、長鎖シーケンサーのデータから遺伝的多様性と体細胞変異を検出する手法を開発し、ヒトの全ゲノムシーケンスを行うことで、ヒトの遺伝的多様性と体細胞変異の全体像の解明を試みた。

ICGC（国際がんゲノムコンソーシアム）（注7）の研究において次世代シーケンサーで解析された11人の肝臓癌と正常組織由来のDNAサンプル（合計22サンプル）の全ゲノ

ムシークエンスを、長鎖シークエンサー（Oxford Nanopore）を用いて行った（サンプルは理化学研究所 中川英刀チームリーダーより提供された）。そしてエラーのパターンを検討し、変異/多型検出手法（CAMPHOR ソフトウェア）を開発した。100bp 以上の挿入・欠失と逆位・転座を対象とし、遺伝的多様性と体細胞変異を検出した（図2）結果、遺伝的多様性解析において 8,004 挿入、6,389 欠失、27 逆位が検出された。挿入と欠失は短いものが多かったが、挿入には、300 塩基付近の長さのもの、6000 塩基付近の長さのものが多い特徴があった（図3）。挿入・欠失の配列の特徴を調べた結果、挿入の 9 割はトランスポゾン由来である一方、トランスポゾンが関係している欠失は少ないことが明らかになった（図4）。また、processed pseudogene（注8）の多型が 15 個検出された。processed pseudogene の起源は、多くの組織で高く発現している遺伝子であることが明らかとなった。肝臓癌の体細胞変異解析では、919 個の体細胞変異が検出された。そのうち、499 個は先行研究で次世代シークエンサーを用いて検出されていたものであり、420 個は新規であった。さらに、構造異常の組み合わせが一部明らかになった。

構造異常の切断点の詳細な解析から、構造異常の原因を推定したところ、非相同組み替え（NAHR）、alternative end-joining（alt-EJ）、FoSTeS/MMBIR、非相同末端結合修復（NHEJ）によって生じたと考えられる構造異常が検出された（図5）。またそれらの割合は、体細胞変異と遺伝的多様性で異なっていた。NAHR による欠失は、短鎖次世代シークエンサーを用いた先行研究では検出されていなかったことから、長鎖シークエンズ技術は従来の研究では見逃されている構造異常を検出するために有用であることが示唆された。

本研究では長鎖シークエンズ技術の解析手法を構築するとともに、ヒトゲノムの多型や変異の全体像の解析を行ったが、今回開発した手法は遺伝性疾患の原因遺伝子探索やがんの変異解析にも利用可能であり、貢献が期待される。

本研究は、日本医療研究開発機構（AMED）「ゲノム医療実現推進プラットフォーム事業」（藤本明洋）、文部科学省科学研究費補助金 基盤 B（課題番号 18H02680 藤本明洋）、文部科学省科学研究費補助金 新学術研究「ゲノム配列を核としたヤポネシア人の起源と成立の解明」（課題番号 18H05511 藤本明洋）、「臨床ゲノム情報統合データベース整備事業」（溝上雅史）の支援を受けて行われた。

5. 発表雑誌：

雑誌名：「Genome Medicine」（2021年4月29日）

論文タイトル：Whole genome sequencing with long-reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer

著者：Akihiro Fujimoto*（東京大学大学院医学系研究科）、Jing Hao Wong（東京大学大学院医学系研究科）、Yukiko Yoshii（京都大学医学研究科）、Shintaro Akiyama（国立長寿医療研究センター）、Azusa Tanaka（東京大学大学院医学系研究科）、Hitomi Yagi（京都大学医学研究科）、Daichi Shigemizu（国立長寿医療研究センター）、Hidewaki Nakagawa（理化学研究所 生命科学センター）、Masashi Mizokami（国立国際医療センター）、Mihoko Shimada（東京大学大学院医学系研究科）

DOI 番号：10.1186/s13073-021-00883-1

アブストラクト URL：<https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-021-00883-1>

6. 問い合わせ先：

東京大学大学院医学系研究科国際保健学専攻

教授 藤本 明洋（ふじもと あきひろ）

電話：03-5841-3692

Mail：afujimoto@m.u-tokyo.ac.jp

7. 用語解説：

注1 長鎖シーケンサー

長い配列を決定できる塩基配列決定装置。Pacific Bioscience 社の PacBio シーケンサーと Oxford Nanopore 社の Nanopore シーケンサーが知られている。長い配列が決定できるため、構造異常の解析に有用であると考えられているが、必要な DNA 量が多いことや塩基配列の解読精度などに課題があることが知られている。

注2 全ゲノムシーケンス

全ゲノムの配列を一度に決定すること。10年ほど前に次世代シーケンサーが登場し、全ゲノムシーケンスが可能になった。しかし、ヒトゲノム内にある多型や変異、構造異常などの全てを検出する方法は十分には確立されていない。

注3 次世代シーケンサー

DNA を断片化し蛍光を用いて塩基配列決定を行う装置。多数の DNA 断片を同時に配列決定するため、大量のデータが得られる。次世代シーケンサーにより、全ゲノムを対象とした遺伝学研究が可能となり、疾患研究に大きな発展をもたらした。エラー率（配列の間違いの割合）が低く、データ産出量も多いものの、読み取り長（リード長）が 100-300 塩基程度と短いため、構造異常の検出において弱点があると考えられている。

注4 構造異常

転座、逆位（DNA 配列の逆転）、挿入、欠失などの総称。ゲノム DNA に大きな変化をもたらすと考えられている。

注5 トランスポゾン

ゲノム内の別の位置に移動する塩基配列。ヒトゲノムの訳 40%がトランスポゾンに由来すると考えられている。また、トランスポゾンの挿入が遺伝子の構造や発現量に影響し疾患の原因となることも報告されている。

注6 非相同組替え

高い相同性を持つゲノム領域の間で誤って生じた組替え。構造異常の原因となると考えられる。非相同組替えによって生じた構造異常には、切断箇所周辺の配列に大きな類似性（100 塩基以上）が存在すると考えられており、切断点周辺の類似性の大きさから非相同組替えによって生じた構造異常か否かが推定される。

注7 国際がんゲノムコンソーシアム (ICGC)

がんにおける遺伝子異常の解明を目的とした国際共同プロジェクト(2008-2018年)。日本、アメリカを含む 16 カ国および EU の機関や組織が参画し、次世代シーケンサーを用いて大規

模なゲノム配列解析が行われた。日本からは、理化学研究所と国立がん研究センターが参加した。本研究では、次世代シーケンサーとの比較解析のため、ICGCのために理化学研究所が収集した DNA サンプルを用いた。

注8 processed pseudogene

転写された mRNA が逆転写されてゲノム DNA に挿入されて生じた偽遺伝子（遺伝子配列は持つが機能を持たないもの）。正常な遺伝子が持つイントロンがないなどの特徴を持つ。

引用文献 1 Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. Cell. 2013;153:919–29.

8. 添付資料：

(1) 構造異常の検出



リードが短いため、ほぼ同じ配列が沢山あると、正しい位置が分からない

(2) 挿入配列の全長



リードが短いため、挿入配列の全長がわからない

図1 次世代シーケンサーの弱点

リード長が短いため、繰り返し配列を介した構造異常の同定が難しい。また、リード長を超える長さの塩基配列の挿入の検出が困難である。

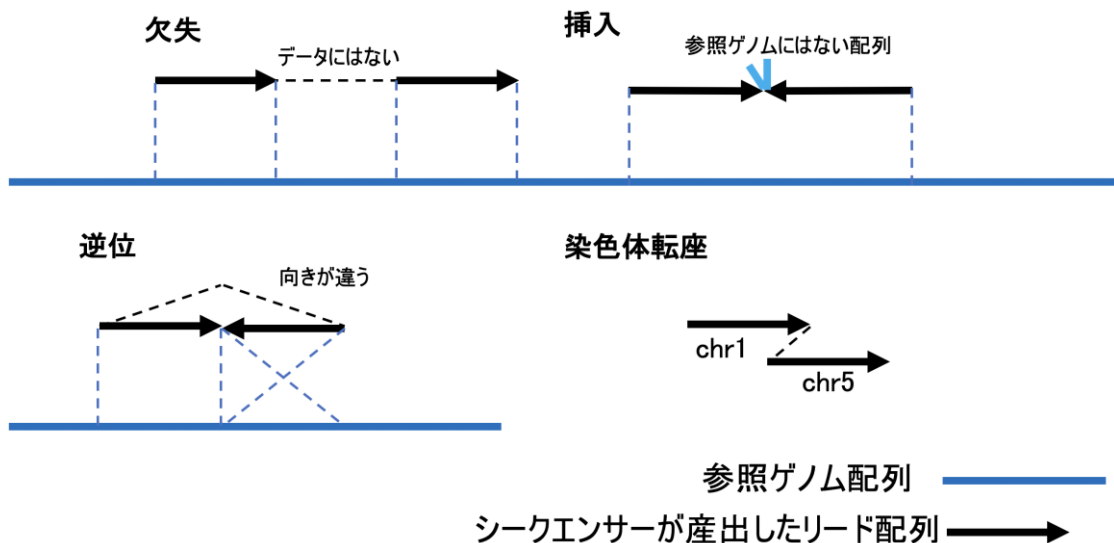


図2 シーケンサーが産出したデータの解析

リード配列と参照ゲノム配列（国際コンソーシアムにより決定され公開されているヒトゲノム配列）との比較により、構造異常を検出した。

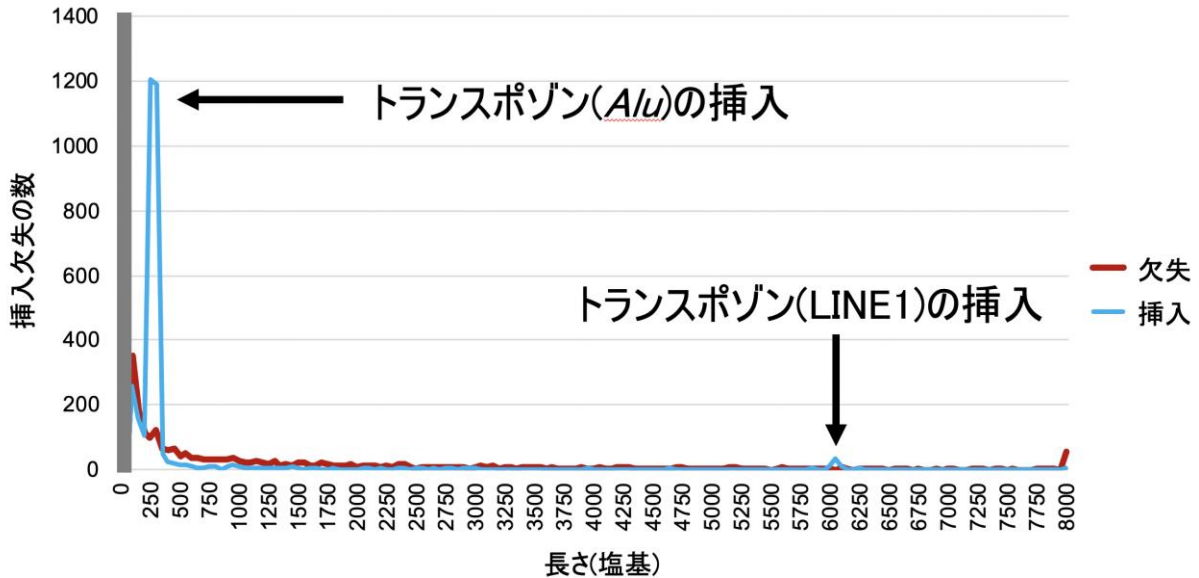


図3 日本人 11 人の全ゲノムシーケンスから同定された挿入欠失の長さの分布
100 塩基未満の挿入欠失は解析対象としていないため、0-100 塩基の領域は灰色で塗りつぶしてある。挿入と欠失で長さのパターンが異なる。挿入には、300bp 付近の長さのもの、6kbp 付近の長さのものが多く、これらはトランスポゾンの挿入であった。

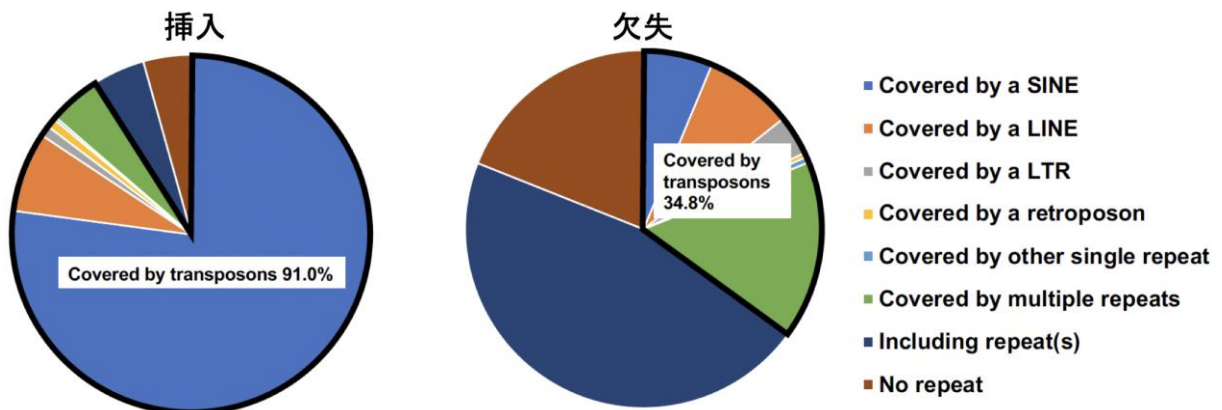


図4 挿入と欠失の配列の特徴

挿入欠失の配列のうち 8 割以上が占めている場合に“覆われている (covered)”とした。挿入は、91%が繰り返し配列に覆われ、84.3%が一種類のトランスポゾン (SINE や LINE) に由来していた。一方、欠失では、34.8%のみが繰り返し配列に覆われており、一種類のトランスポゾン (SINE や LINE) に覆われた領域の欠失は少なかった (SINE は 6.3%、LINE は 8.1%)。

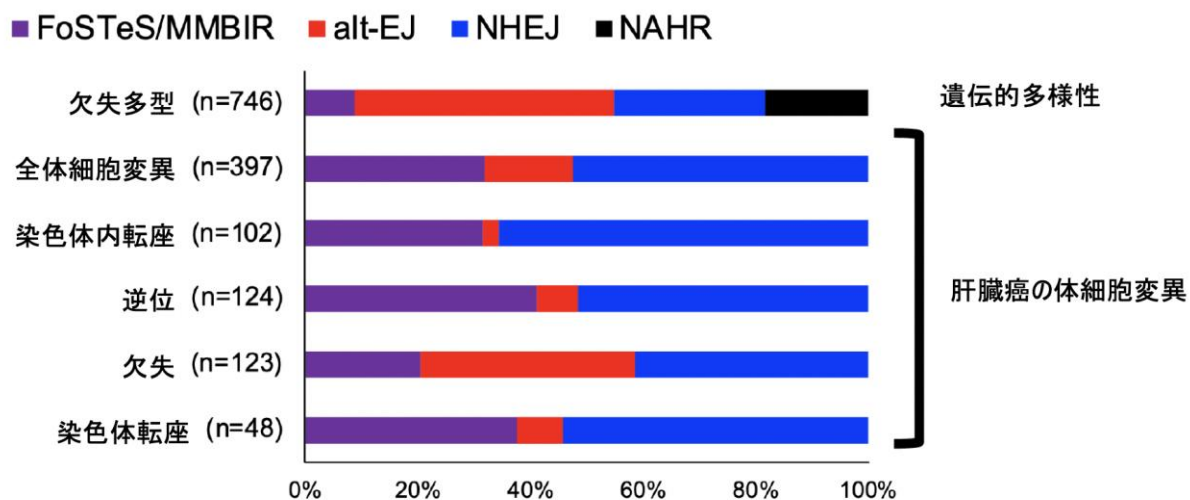


図5 構造異常の起源の推定

遺伝的多様性と体細胞変異を比較すると、NAHR（非相同組み替え）による構造異常は、遺伝的多様性にもみ見られた。生殖細胞での変異を起源とする遺伝的多様性と体細胞変異では、構造異常の原因に違いがあることが示唆された。