

## 科学研究費助成事業 研究成果報告書

令和 5 年 6 月 17 日現在

機関番号：82729

研究種目：若手研究

研究期間：2020～2022

課題番号：20K16945

研究課題名（和文）ロングリードシーケンサーによる染色体構造異常症の切断点解析

研究課題名（英文）Breakpoint analysis of balanced chromosomal abnormalities by long-read sequencer

研究代表者

村上 博昭（Murakami, Hiroaki）

地方独立行政法人神奈川県立病院機構神奈川県立こども医療センター（臨床研究所）・臨床研究所・医師

研究者番号：30836440

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：遺伝情報の欠失や重複を伴わない均衡型染色体構造異常では、切断点が疾患に関連する遺伝子を直接もしくは間接的に破壊することが疾患の原因となると考えられる。しかし、既存の遺伝学的解析方法では、切断点を詳細に解析することが困難であるため、未診断症例が多数存在する。本研究は、ロングリードシーケンサーによる均衡型構造異常の解析の臨床的有用性を評価することを目的とした。計4例の症例について解析を行い、うち3例で詳細な切断点を同定し、先天異常の遺伝学的原因を同定することができた。均衡型染色体構造異常の遺伝学的解析に、ロングリードシーケンサーが非常に有用であることを示すことができた。

研究成果の学術的意義や社会的意義

本研究では、4例の均衡型染色体構造異常を有する先天異常症例について、ロングリードシーケンサーによる切断点解析を行った。うち3例で詳細な切断点を同定し、先天異常の原因と考えられる遺伝学的な要因を同定できた。今後、同手法が未診断症例の診断率向上に寄与する可能性が非常に高く、社会的な意義は大きいと考えられた。また、切断点が原因遺伝子を直接切断するのではなく、染色体高次構造を破壊することが疾患発症メカニズムとなる可能性も示唆され、学術的にも非常に興味深い結果が得られた。

研究成果の概要（英文）：The balanced chromosomal abnormalities, which do not exhibit apparent deletions or duplications, are challenging to analyze using traditional genomic technologies. The precise identification of breakpoints is difficult, resulting in many undiagnosed patients with balanced chromosomal abnormalities. In this study, we explored the clinical utility of long-read sequencing analysis for such abnormalities. Four cases were analyzed, and genetic causative genes could be identified in two of them. This study highlights the clinical benefits of long-read sequencers in identifying balanced chromosomal structural variants.

研究分野：臨床遺伝学

キーワード：均衡型染色体構造異常 切断点解析 ロングリードシーケンサー 先天異常症

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

先天異常の原因診断のために行われる染色体検査において、転座や逆位は遺伝情報の明らかな量的変化(コピー数異常)を伴わないため、均衡型染色体異常(balanced chromosomal abnormalities: BCAs)と呼ばれる。一般人口における BCAs の頻度が 0.2-0.5%であるのに対し、知的障害(1.5%)、発達障害(1.3%)では頻度が増加し、その多くが両親にはない新生の変化(*de novo*)であることが知られている。また、*de novo* BCAs を有する胎児の出生後の先天異常の頻度は、転座と逆位を合わせて 6.4%と報告されている(Warburton, 1991)。さらに近年の報告では、詳細に解析された BCAs のうち、30%もの BCAs の切断点が疾患関連遺伝子を破壊し、7.3%は染色体機能単位としての高次構造である TADs (Topologically associated domains) を破壊することが示された。BCAs はこれまでの予測以上に疾患の原因となっている可能性が示唆された(Redin, 2017)。BCAs の解析には、切断点の詳細を明らかにする必要があるが、マイクロアレイ染色体検査やショートリードシーケンサー(リード長=100-200bp)を用いた従来の解析では切断点の解析は極めて困難であり、実臨床においても *de novo* BCAs を有する未診断の先天異常患者は多数存在している。近年実用化されたロングリードシーケンサーは新しい原理による DNA の塩基配列解析技術であり、リード長は平均で 10-20kb にもなる。そのため、1本のリードで切断点をまたいで広範囲の配列が解析可能であり、転座や逆位の切断点の解析に有用である可能性が示唆されている。

### 2. 研究の目的

本研究の目的は以下の通りである。

ロングリードシーケンサーが BCAs の切断点の解析に有用であるかどうかを評価すること

ロングリードシーケンサーによる BCAs の切断点の解析が、未診断の先天異常患者の診断率向上や健康管理の改善に寄与するかどうかを評価すること(臨床上の意義)

染色体異常による先天異常発症メカニズムについて、新たな学術的知見が得られるかどうかを評価すること

### 3. 研究の方法

研究代表者の外来には、年間約 450 人の新規の先天異常患者が受診する。またこれまでの蓄積として、10,000 人以上の染色体検査結果、1,300 人以上のマイクロアレイ染色体検査結果、1,000 人以上のショートリード解析データを保有している。本研究の解析は以下のように進めた。

未診断の先天異常患者のうち、BCAs を有する患者(及び両親)に研究内容を説明し、書面での同意を得る。その後、採血を行い末梢血 DNA を抽出する。

Oxford Nanopore 社の Ligation Sequencing Kit 1D を用いて DNA ライブラリを作成した後、同社の Nanopore PromethION sequencer によりシーケンスを行う。

得られたシーケンスデータをヒトゲノム参照配列(hg19)に対して mapping を行う。

染色体構造異常の検出とアノテーションは、それぞれ Sniffles v1.0.11 (Sedlazeck, 2018)、Snpeff v4.3q (Cingolani, 2012) を用いて行う。

検出した構造異常を Integrative Genomics Viewer や Ribbon (Nattestad, 2021) を用いて可視化し、切断点の同定を行う。

疾患関連遺伝子の破壊や TAD の破壊などが予想される場合は、必要に応じて RNA-seq などを用いた転写 RNA の評価を行う。TAD の評価はオープンソースの TAD database を適宜利用する。

解析結果と臨床像の整合性については、専門他科(小児放射線科、小児神経内科など)を含めたカンファレンスにおいて議論し、最終的な遺伝学的診断を行う。

### 4. 研究成果

(解析を施行した 4 症例の解析結果)

本研究目的に一致する以下の 4 症例を解析対象とした。全て染色体検査により各種の BCA を有することが分かっていたが未診断の先天異常疾患患者であった(表 1)。患者末梢血 3-5ml を用いて、DNA 3µg 以上(Qubit による測定)、精製純度(OD 260/280 =1.8-2.0、OD 260/230 =2.0-2.2)、電気泳動による平均分子長 30kb などの Nanopore ロングリードシーケンスのサンプル要件が確保できた。シーケンスから得られたデータ量は 81.2-111.7Gb、総 Read 数は 11.9-13.6x10<sup>6</sup>、平均 Read 長はおよそ 9-10kb であった。

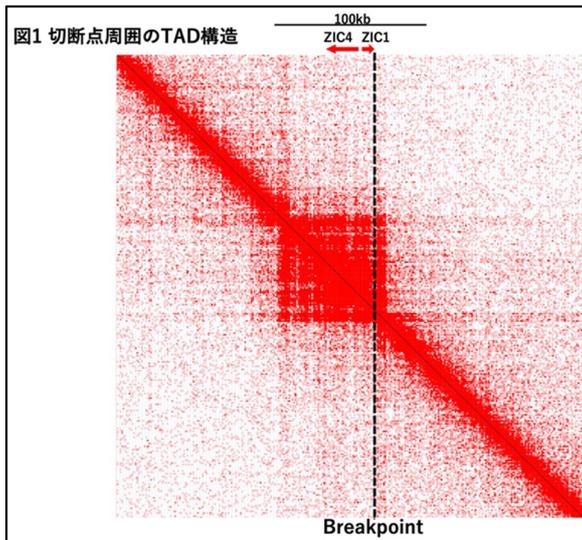
以下、解析対象とした 4 症例について、個人情報に配慮した

表 1 解析症例まとめ

症例	BCA の分類	染色体検査結果	マイクロアレイ	全エクソーム解析	<i>de novo</i>
1	逆位	46,XY,inv(3)(○○)	異常なし	異常なし	○
2	転座	46,XY,t(Y;1)(○;○)	異常なし	異常なし	○
3	転座	46,XX,t(9;17;20)(○;○;○)	異常なし	未施行	未確認
4	逆位	46,XY,inv(3)(○○)	未施行	未施行	○

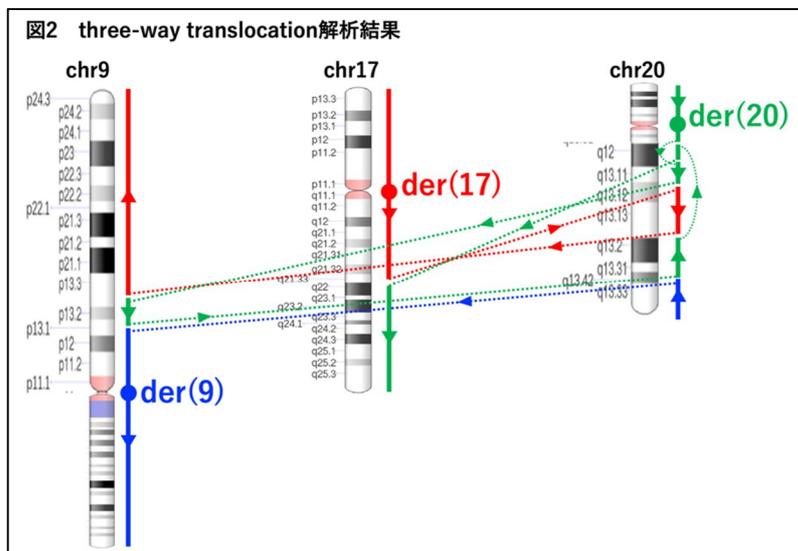
形で臨床経過を示す。

(症例1) 男児。前額突出、耳介低位などの特徴的顔貌、脳形成異常(水頭症、菱形シナプス、Chiari 奇形など) 頭蓋骨早期癒合、後頭部・側頭部の禿頭など多発奇形に加え、精神運動発達遅滞を示した。臨床診断は Gomez-Lopez-Hernandez 症候群 (GLHS、Sukhudyen, 2010) であったが、GLHS は原因遺伝子不明の疾患であり、遺伝学的には確定診断ができなかった。逆位の範囲は 38Mb と大きく 500 個近い遺伝子を含んでいた。ロングリードシーケンスの結果、3 番染色体に 9.7Mb の逆位を検出した。2 か所の切断点を解析したところ、共に遺伝子のコーディング領域には含まれなかったが、テロメア側の切断点は転写因子である ZIC1 の 3'-非翻訳領域 (UTR) の 7kb 下流に位置していた。ZIC1 は小脳を中心に脳特異的に発現する転写因子であり、zinc finger ドメインを有する。過去に報告は少ないものの、機能喪失変異では Dandy-Walker 奇形、機能獲得変異では特徴的顔貌、頭蓋骨早期癒合、脳奇形、発達遅滞を示すことが示されており、本患者の臨床像に極めて一致していた (Grinberg, 2004、Twigg, 2015)。TAD データベースの解析では、切断点が ZIC1 と ZIC4 を含む TAD を破壊している可能性も示唆された (図 1)。ZIC1 の異常転写産物の検出を目的として、患者と両親の末梢血を用いた RNA-seq 解析を施行したが、末梢血リンパ球では ZIC1 の転写産物は検出できなかった。



(症例2) 男児。成長障害、心奇形を含む多発奇形、精神運動発達遅滞を示していた。1 番染色体と Y 染色体に *de novo* の均衡型転座を検出していたが、マイクロアレイ染色体検査 (CMA) や全エクソーム解析 (WES) では疾患関連のコピー数異常や塩基変化は検出できず、遺伝学的診断には至らなかった。ロングリードシーケンスの結果、十分なデータ量 (105.8Gb) とリード長 ( $11.9 \times 10^6$  reads、平均長 9.8kb、マップ率 95.8%) を得られたが児の均衡型転座の切断点を同定することができなかった。

(症例3) 女児。先天性水頭症、食道閉鎖、斜視、右母指低形成などの多発奇形に加え、精神発達遅滞、てんかんを認めた。染色体検査では 9 番、17 番、20 番染色体の均衡型 three-way translocation が検出されたが、CMA では明らかな疾患関連となるコピー数異常は検出されなかった。ロングリードシーケンスの結果、染色体検査結果とは大きく異なる非常に複雑な染色体構造異常であることが分かった (図 2)。染色体の切断点は 9、17、20 番染色体でそれぞれ 3 か所、2 か所、7 か所検出され、微細欠失領域もそれぞれ 1 か所、1 か所、2 か所検出された。特に der(20) は 3 本の染色体それぞれの領域を含む複雑な構造で構成されていることが明らかになった。検出された切断点や欠失領域を詳細に検討した結果、20 番染色体の切断点の一つがハプロ不全により発達遅滞の原因となることが報告されている PTPRT 遺伝子を切断していることが分かった。PTPRT は脳に特異的に発現するチロシンキナーゼをコードする。



(症例4) 男児。先天性心疾患、停留精巣、精神運動発達遅滞、重度成長障害を有し、染色体検

査で両親には認めない、*de novo*の3番染色体逆位が検出された。ロングリードシークエンスの結果、当初予想されていた3番染色体の逆位に加えて、4番染色体と5番染色体の長腕末端の不均衡型相互転座と12番染色体の腕内逆位が検出された。児の先天異常の表現型は4番染色体長腕の欠失と5番染色体長腕の重複が原因であった。また、3番染色体の逆位の切断点の一方は、神経発達に關与するグルタミン酸受容体の一つであるGRM7を切断していた。GRM7はハプロ不全を予測するスコアが非常に高い( $pLI=1.0$ )が、過去の症例はホモ接合体のミスセンス変異症例がほとんどであり本児の表現型への關与については現時点で不明であった。

#### (本研究の考察)

本研究では、BCAsを有する未診断の先天異常患者4例に対してNanoporeシークエンサーを用いたロングリード解析を施行した。全ての解析でゲノム全長(3.2Gb)をカバーする十分なデータ量(81.2-111.7Gb)を得ることができ、マップ率は95%以上が確保できた。4症例のうち3症例(75%)で、詳細な切断点を同定することができた。切断点が同定できた3症例については、いずれも患者の先天異常の原因、もしくは原因の一部となっている遺伝子を同定することができた。唯一切断点が同定できなかった症例2は、転座切断点の一方がY染色体に存在することが影響している可能性があると考えられた。1コピーしか存在しないY染色体は、他の常染色体と異なりリード数(read depth)が少なくなる傾向がある。また繰り返し配列に富むヘテロクロマチン領域が多いため、マップ率も低下する。ロングリードでの解析であっても、このように繰り返し配列に富む領域においては、解析成功率が下がる可能性が示唆された。今後、このような対象においても解析成功率を上げるための工夫が必要であると思われる。

症例1では、切断点が原因遺伝子ZIC1のコード領域外(3'-UTRの7kb下流)に存在した。表現型からは、ZIC1の機能獲得型の変化が生じていると考えられるが、具体的な分子メカニズムは明確にできなかった。可能性としては、TADの破壊による転写調節異常、C末端が異常伸長した活性型ZIC1の出現などが考えられた。を支持するデータとしては、共に脳特異的に発現するzinc-fingerファミリーの転写因子であるZIC1とZIC4が染色体上に隣接して存在しており、TADデータベースで両者を含むTADが確認できることが挙げられる。胎児期の特定の時期や部位において正確に制御されるべきZIC1とZIC4の転写が、TADの破壊により異常を生じたのかもしれない。については、RNA-seqによる異常転写産物の検討が有効であると考えられたが、残念ながら患者由来の末梢血リンパ球では、ZIC1、ZIC4ともに転写産物の検出ができなかった。このように、切断点によるTADの破壊などの転写調節異常の分子レベルでのメカニズムの解明は、胎児期の器官形成期特有の異常が原因となっている可能性が高く、末梢血の解析のみでは評価が困難であることが、今後も課題であると思われる。これらの評価のためには、iPS細胞やモデル生物を用いた基礎研究を行う必要がある。

症例3では、three-way translocationの症例の解析の結果、染色体検査では全く予想できなかった複雑な転座や欠失を明らかにすることができ、その結果、症状の一部を説明する原因遺伝子を同定することも可能であった。このような高度な複雑染色体構造の解析には、1本のリードで多数の切断点をまたぐことが可能なロングリードシークエンスが非常に有用であった。ただし、一方で、IGVやRibbonで可視化したデータを目視で確認して全体構造を把握する作業は膨大な時間を必要とした。より複雑な構造異常を対象とする場合には、自動で構造解析する専用のプログラムの利用が必須であり、そのためには高度なデータ解析環境が必要と考えられた。

本研究では、先天異常の患者における均衡型染色体構造異常の解析においてロングリードシークエンサーが非常に有用な解析ツールとなることを示すことができた。今後、未診断の先天異常患者の診断率の向上に寄与する可能性が非常に高く、臨床的なインパクトは大きい。また、新規の原因遺伝子の発見や新たな疾患メカニズムの解明につながる可能性もあり学術的なインパクトも大きいと考えられた。ただし、切断点が必ずしも原因遺伝子を直接破壊しているとも限らず、その場合には、確定診断のためにより詳細な転写産物の解析や基礎研究が必要となる場合もあることが分かった。今後も染色体構造解析においてロングリードシークエンサーは臨床的、学術的に非常に有用なツールとなると結論する。

#### (引用文献)

1. Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at perinatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet.* 1991 49(5):995-1013.
2. Redin C, Brand H et al. The genomic landscape of balanced cytogenetic abnormalities

associated with human congenital anomalies. *Nat Genet.* 2017 49:36-45.

3. Sedlazeck FJ, Rescheneder P et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods.* 2018 30;15(6):461-8.

4. Cingolani P, Platts A et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012;6(2):80-92.

5. Nattestad M, Aboukhalil R, Chin CS, Schatz MC. Ribbon: intuitive visualization for complex genomic variation. *Bioinformatics.* 2021;37(3):413-5

6. Sukhudyana B, Jaladyan V et al. Gómez-López-Hernández syndrome: Reappraisal of the diagnostic criteria. *Eur J Pediatr.* 2010;169(12):1523-8.

*Bioinformatics.* 2021;37(3):413-5

7. Grinberg I, Northrup H et al. Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat Genet.* 2004;36(10):1053-5.

8. Twigg SRF, Forecki J et al. Gain-of-Function Mutations in ZIC1 Are Associated with Coronal Craniosynostosis and Learning Disability. *Am J Hum Genet.* 2015;97(3):378-88.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 村上博昭、榎本友美、熊木達郎、黒澤健司
2. 発表標題 De novoの3q腕内逆位を有し、Gomez-Lopez-Hernandez症候群を疑った男児におけるナノボア長鎖シーケンサーによる切断点解析
3. 学会等名 第67回 人類遺伝学会
4. 発表年 2022年

1. 発表者名 村上博昭、熊木達郎、榎本友美、阪下達哉、寺澤篤志、今村淳、黒澤健司、金子英雄
2. 発表標題 STAG1を含むde novoの3q22.1q22.3欠失を有する、重度成長発達遅滞、多発奇形の男児例
3. 学会等名 第67回 人類遺伝学会
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------