

令和 4 年 6 月 26 日現在

機関番号：82612

研究種目：基盤研究(C) (一般)

研究期間：2017～2021

課題番号：17K08690

研究課題名(和文) サブテロメア関連疾患の原因究明に向けたゲノム情報基盤の構築

研究課題名(英文) Genome sequence analysis of the subtelomeric regions to understand the etiology of subtelomere-related syndromes.

研究代表者

黒木 陽子 (Kuroki, Yoko)

国立研究開発法人国立成育医療研究センター・ゲノム医療研究部・室長

研究者番号：10344037

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、ヒトのサブテロメア領域のゲノム解析を実施し、サブテロメアにおけるゲノム構造と疾患との関連性及びゲノム多様性の解明を目指して研究を実施した。染色体特異的ゲノムライブラリーから単離したヒトサブテロメア領域由来のクローン40個について、ロングリードシーケンサーによる配列解析を実施した。上記の解析から得られたデータのデータ統合においては、ショートリードシーケンサーによる配列補正を行い、高精度なクローン毎の配列データの構築を進めている。また、サブテロメア領域のゲノム多様性の実施に向けた、既報のゲノムデータセットの収集を進めている。

研究成果の学術的意義や社会的意義

本研究では、ヒトゲノムの難解読領域の一つであるサブテロメア領域を対象として、高精度なゲノム配列データの取得及びこの領域における疾患との関連性を明らかにすることを目指して解析を行なった。サブテロメア領域は、多くの反復配列が存在することから、これまで十分な解析が行われず、ゲノム多様性や疾患との関連性は明らかになっていない。本研究で得られる配列データと既に公開されているヒトゲノム参照配列の比較解析をすることにより、サブテロメア領域のゲノム多様性の一端が明らかになる。また、ヒトのサブテロメア領域におけるゲノム多様性の解明は、将来的にサブテロメア関連疾患の原因究明に繋がることが期待される。

研究成果の概要(英文)：We carried out the genome analysis of the human subtelomeric regions. The purpose of this study is to elucidate the etiology of the subtelomere-related syndromes and to examine the diversity of subtelomere genomic structures. The subtelomeric clones which have been obtained by PCR-based screening from human chromosome-specific genome libraries, and these clones were sequenced for both ends of the inserted sequence. Then, the forty clones which located in subtelomeric regions and seems to have novel sequence were analyzed by the long read sequencer. Although obtained sequences were assembled for each clone, it was unable to construct a precise and a high-quality data by itself. Therefore, the short read sequence data were added to the long read data, the precise and single contig sequence were constructed for each clone. To identify genome diversity, we will examine the sequence similarity to the assembled data against the several genome reference data set which are already published.

研究分野：ゲノム科学

キーワード：サブテロメア ゲノム難解読領域 ロングリードシーケンサー 次世代シーケンサー 散財性反復配列 サブテロメア欠失症候群 多発奇形 精神遅滞

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

サブテロメア領域は、染色体の末端に位置するテロメア領域と各染色体に特異的なゲノム領域との間の境界部分で、SINE や LINE などの散在性反復配列に加え、サテライトリピートやテロメア様配列が密集する、特徴的なゲノム構造を示す¹⁾。この領域は、他の染色体のサブテロメア領域と配列相関性が高く、ゲノム全領域の配列データを用いた解析では、個々のサブテロメア領域を識別することが難しいため、詳細な構造を明らかにすることは困難である。申請者らは、これまでの研究において、独自に構造した染色体特異的ゲノムライブラリーを用いて、各染色体のサブテロメア領域由来のクローンを単離し、クローンの末端配列決定を行った。この方法を用いて、205 個のサブテロメア候補クローンを単離し、そのうちの 25 クローンについてはクローンの内部配列を決定し、得られた配列データをデータベースに登録した。それらの配列データは、ヒトゲノムの参照配列 (NCBI GRCh38/UCSC hg38) に組み込まれている。一方、残りの 180 クローンについては、反復配列の長さや構造上の複雑さ故に、従来法である Sanger 法では完全配列の取得が難しく、未解読のまま残されている。ヒトゲノム計画は 2003 年に終了し、この研究により得られたデータを基盤として構築されたヒトゲノムの参照配列は、現在もその情報が更新され続けているが、サブテロメア領域近傍は、未だクローンギャップやシークエンスギャップが存在している状況である。このように、サブテロメア領域は、解析の基盤となる参照配列の精度が十分でないため、集団におけるゲノム多様性解析、疾患関連解析が立ち遅れているのが現状である。

一方、サブテロメア領域は、原因不明の様々な疾患と関連することが報告されている。サブテロメア領域の微細なゲノム構造異常 (欠失、重複、あるいは逆位) を原因とする先天異常症候群は、構造異常が見られる領域によって、1p36 欠失症候群、6p25 欠失症候群、9q34 欠失症候群、22q11.2 欠失症候群などが知られている^{2), 3)}。現状では、マイクロアレイを用いた方法⁴⁾や染色体 FISH 法により、サブテロメア領域の構造解析を行なっているが、ヒトゲノム参照配列が不十分であることや、一般集団におけるゲノム多様性が未解明であることから、疾患と直接関わる構造異常の同定が難しい。また、異常が検出された場合でも、切断点の同定や構造変化の発生機序は殆ど明らかになっていない。ヒトサブテロメア領域は、その生物学的意義やゲノム機能について未だ十分な解析が行われていない。

2. 研究の目的

本研究では、サブテロメア領域のゲノム情報基盤の整備と、この領域のゲノム多様性や疾患との関連を明らかにする。また、健常人と患者 (サブテロメア微細構造異常症や多発奇形、精神発達遅滞の症状を呈す患者) を対象としたゲノム解析手法の開発と、疾患の原因遺伝子の同定や染色体構造異常の発生機序の解明を目指して研究を進める。多発奇形や精神発達遅滞の症状を示す患者の約 10% は、サブテロメア領域のゲノム構造異常に起因すると考えられている。サブテロメア領域は、ヒトゲノムにおいて解析が難しい領域として知られており、ヒトゲノム参照配列のデータを見ても、配列データの精度が不十分な上に、未解読な領域が多く存在する。本研究では、ヒトサブテロメア領域のゲノム構造解析を行い、ゲノム情報を整備する。健常人における多様性を明らかにし、この領域に特化した解析手法の開発を行う。また、多発奇形や精神発達遅滞の患者を対象としたゲノム解析を行い、疾患の原因遺伝子の探索と患者で見られるゲノム構造異常の発生機序の解析を目指す。

3. 研究の方法

本研究では、申請者らが作成した、健常人由来の染色体特異的なサブテロメアクローンを研究材料として、ロングリードシークエンサーである Pacific Bioscience 社の RSII を用いたゲノム構造解析を行った。得られた配列データをヒトゲノム参照配列データと比較し、参照配列でギャップとなっているゲノム領域を補完するとともに、サブテロメア領域のゲノム配列情報の高精度化を目指して研究を行った。

1) 新規ゲノム配列を含むサブテロメアクローンの選別

染色体特異的ゲノムライブラリーから単離したサブテロメア候補クローンは、Sanger 法による末端塩基配列を決定し、最新版のヒトゲノム参照配列である GRCh38/hg38 の配列データとの比較解析を実施した。この解析により、GRCh38/hg38 において未解析の領域とオーバーラップするクローンを選別し、ロングリードシークエンサーによる配列解析を行った。GRCh38/hg38 において未解析となっている領域の多くは、クローンギャップまたはシークエンスギャップとなっており、それらの領域と相関性を示す配列を含むサブテロメアクローンは、新規ゲノム配列を含んでいる可能性が高い。そこで、180 の候補クローンのうち、末端塩基配列が、GRCh38/hg38 の配列データでギャップになっている部分の近傍に相関性を示し、クローンの内部配列がギャップ領域を含んでいるクローンを優先的に、以降の解析を行なった。

2) ロングリードシークエンサーによる配列解析

1) の解析において選別したクローンについて、Pacific Biosciences 社、RSII による配列データの取得を行った。RSII による配列解析では、1 SMRT cell あたりのクローン数を 12 クローンとして、各クローンに特異的なバーコードアダプターを付与し、multiplex sequencing を行った。また、得られた配列データは、バーコードアダプターによる各リードの分類を行い、クローン毎の配列データセットの構築を行った。

4. 研究成果

本研究は、サブテロメア領域のゲノム構造解析を行い、ゲノムの構造や塩基配列の多様性を明らかにするとともに、ゲノム構造と疾患との関連解明を目的として研究を実施した。我々は、過去に実施したヒトゲノム研究から、サブテロメア領域に位置すると考えられる Fosmid クローンを、180 個保有していた。これらのクローンのうち、末端配列の相同性解析から、サブテロメア領域由来のクローンで新規のゲノム配列を含むことが予想された 40 クローンについて、ロングリードシーケンサーによる配列解析を実施した。ロングリードシーケンサーにより得られた配列データは、SMRT Link による配列統合を試みたが、クローン毎の配列データの統合に苦労している。データ統合が難しい理由として、サブテロメア領域に位置する様々なタイプの反復配列が、ロングリードシーケンサーから得られた配列データのみで環状化(1本の配列データにすること)することが困難であった。そこで、ロングリードシーケンサーのデータを補正するため、illumina 社の MiSeq による配列データを取得し、ロングリードシーケンサーのデータとショートリードシーケンサーのデータの統合解析(ハイブリッドアセンブリ)を試みている。これらの解析から、クローン毎に、約 40 kb の精度の高い配列データの構築を進めている。

一方、2022 年 4 月に、Telomere-to-Telomere Consortium による T2T-CHM13 project による欧米人由来のゲノム参照配列が公開された⁵⁾。この参照配列は、ヒトの一倍体ゲノム(常染色体 22 本と X 染色体)が倍化した細胞由来の全ゲノム配列データセットである。具体的には、受精時の発生異常で生じる胞状奇胎と呼ばれる疾患由来の細胞で、1 倍体の染色体が倍化して 2 倍体になった細胞のゲノム配列データである。T2T-CHM13 は、シーケンシングギャップが存在しない、全てのサブテロメア領域の配列データを含むと報告されている。今後、T2T-CHM13 を含む、以下のヒトゲノム参照配列、GRCh38/hg38、東北メディカル・メガバンク機構が構築する日本人由来の配列データセットを用いて、本研究で得られた配列データとの比較解析を実施する。これらの解析を通じて、日本人のゲノム解析に有用なサブテロメア領域のゲノムデータ基盤を構築する。

<引用文献>

1. Linardopoulou et al., *Nature*, 2005, 437, 94-100
2. Tsuyusaki et al., *Pediatr Int*, 2010, 52(4), 547-550
3. Yamamoto et al., *Am J Med Genet A*, 2006, 140(12), 1302-1304
4. Harada et al., *J Med Genet*, 2004, 41(2), 130-136
5. Nurk et al., *Science*, 376(6588), 44-53

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計3件（うち招待講演 1件 / うち国際学会 2件）

1. 発表者名 黒木陽子・豊田 敦・野口英樹・谷口丈晃・矢田哲士・小野千紘・要 匡・藤山秋佐夫
2. 発表標題 ヒトサブテロミア領域のゲノム構造、その機能と進化
3. 学会等名 ConBio2017, 2017年度生命科学系学会合同年次大会（招待講演）
4. 発表年 2017年

1. 発表者名 Yoko Kuroki, Atsushi Toyoda, Hideki Noguchi, Chihiro Ono, Asao Fujiyama, Tadashi Kaname
2. 発表標題 Resequencing and analysis of the human subtelomeric regions.
3. 学会等名 AGBT2018, Advances in Genome Biology and Technology2018（国際学会）
4. 発表年 2018年

1. 発表者名 Yoko Kuroki, Atsushi Toyoda, Hideki, Chihiro Ono, Tadashi Kaname, Asao Fujiyama
2. 発表標題 Genome analyses with the long-read sequencing technology toward the construction of high-quality genomic data for the primate subtelomeric regions.
3. 学会等名 International society for computational biology（国際学会）
4. 発表年 2018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
研究協力者	豊田 敦 (Toyoda Atsushi)		

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
---------	---------