

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 10 日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2011～2013

課題番号：23659180

研究課題名(和文) シークエンスレベルでのヒトゲノム構造多様性の解明

研究課題名(英文) Identification and elucidation of structural variations in human genome

研究代表者

馬淵 昭彦 (MABUCHI, AKIHIKO)

東京大学・医学(系)研究科(研究院)・准教授

研究者番号：80312312

交付決定額(研究期間全体)：(直接経費) 2,800,000円、(間接経費) 840,000円

研究成果の概要(和文)：今日のヒトゲノム研究の進展に伴い、特定の形質の原因変異・感受性多型同定法が確立しつつある。特に疾患を対象としたゲノムワイド関連解析研究では、さまざまな疾患関連多型が同定されている。云うまでもなく最多の多型はSNPであるため、現在までに報告されたこれらの疾患関連多型もその多くはSNPである。しかしながら、ゲノム中にはSNP以外にもCNVをはじめとして多種の多型・変異が存在する。本研究では、ヒトゲノム中に存在する均衡型構造多型・変異のうち特に逆位と染色体内・染色体外転座を同定する手法を確立することを目指した。

研究成果の概要(英文)：With the progress of recent studies in human genetics, methods to identify causing or susceptible variations have been established. Especially, many genes and polymorphisms which associated with diseases or phenotypes have been identified. Much of those studies were focused on single nucleotides polymorphisms, which are most frequent variation in human genome. However, other type of variations including copy number variations might associate with phenotypes. Here we focus on the structural variation, especially inversions and translocations and aimed to identify those variations.

研究分野：医歯薬学

科研費の分科・細目：基礎医学・人類遺伝学

キーワード：ゲノム医学

## 1. 研究開始当初の背景

今日のヒトゲノム研究の進展に伴い、特定の形質の原因変異・感受性多型同定法が確立しつつある。特に疾患を対象としたゲノムワイド関連解析研究では、さまざまな疾患関連多型が同定されている。云うまでもなく最多の多型は SNP であるため、現在までに報告されたこれらの疾患関連多型もその多くは SNP である。しかしながら、ゲノム中には SNP 以外にも CNV をはじめとして多種の多型・変異が存在する。

DNA チップの解析から、2006 年 CNV 地図の概要が示され(Nature, 444:444-, 2006)、2010 年 5 月既に 57,892 の CNV、1,4478 の CNV 領域が報告された(DGV: DataBase of Genomic Variants, URL; <http://projects.tcag.ca/variation/>)。SNP 以外の多型のうち CNV 解析が先行した理由としては、DNA チップではシグナル強度を解析することでコピー数の量的解析が可能であること、Affymetrix 社に代表される CNV 専用のプローブが搭載された DNA チップが開発されたこと、また array GCH (aCGH) など既存技術の研究基盤が存在したことが大きい。

一方、DNA の量的変化を伴わない逆位等の均衡型構造多型・変異は、DNA チップでは把握が困難である。CGH によるゲノム構造多型・変異の解析では、現在まで 10 余の報告があるが多型・変異の大多数は CNV、In-del であり、逆位の同定数はごく少数にとどまる。既存技術を用いた逆位の検出で最も確実な方法は、次世代シーケンズ法による全ゲノムシーケンズと考えられる。ごく最近に報告された日本人の全ゲノム配列の包括的解析では、1 個人のゲノムから 57 の逆位と 112 の染色体内転座が示された(Nature genet, advance online publication)。しかしながら、全ゲノムシーケンズは、その費用的な制約によりゲノム構造多型・変異の解析において理想的な手法とは言い難い。

## 2. 研究の目的

ヒトゲノムプロジェクト、HapMap プロジェクトにより、ヒトゲノムに存在する SNP が数多く同定され、また、ヒトゲノム構造多型コンソーシアムをはじめとした多くの国際共同研究により、CNV (コピー数多型) が同定されてきた。しかしながら、その多くはゲノム上の大まかな領域の検出にとどまり、CNV の存在の実験レベルの確認や具体的な切断点の情報に欠く等の問題がある。また、均衡型構造変化である逆位は、DNA チップによる検出が困難でありまたその詳細な地図は存在していないのが現状である。本研究では、特に逆位、染色体内・染色体外転座等の均衡型構造多型・変異が生じている部分の網羅的検出とその切断点の同定を目指して、その方法論を確立すること、及び日本人検体を用いてその均衡型構造多型・変異地図を作

成することを目的とする。

本研究では、ヒトゲノム中に存在する均衡型構造多型・変異のうち特に逆位と染色体内・染色体外転座を同定する手法を確立することを目指す。複数検体をメイトペアシーケンズ法で次世代シーケンサーにより同時にリードすることで費用低減を図りつつ、新規開発する専用プログラムで高速に均衡型構造多型・変異の候補領域を網羅的に検出する。これと平行し逆位領域の切断点同定のため、PCR 法、LAMP 法、MIP 法その他により最も簡便かつ高速に検出できる方法を検討する。検出された候補領域に対して切断点を同定可能か検討し、可能であれば複数領域に対して実施することで均衡型構造多型・変異地図を作成する。更に、申請者らが研究を行っている疾患研究に関してこの多型の関連解析を行う。

そこで本研究では、ヒトゲノム中に存在する均衡型構造多型・変異のうち特に逆位と染色体内・染色体外転座を同定する手法を確立することを目指す。複数検体をメイトペアシーケンズ法で次世代シーケンサーにより同時にリードすることで費用低減を図りつつ、新規開発する専用プログラムで高速に均衡型構造多型・変異の候補領域を網羅的に検出する。これと平行し逆位領域の切断点同定のため、PCR 法、LAMP 法、MIP 法その他により最も簡便かつ高速に検出できる方法を検討する。検出された候補領域に対して切断点を同定可能か検討し、可能であれば複数領域に対して実施することで均衡型構造多型・変異地図を作成する。更に、申請者らが研究を行っている変形性関節症に関してこの多型の関連解析を行い、変形性関節症感受性逆位の同定を目指す。

## 3. 研究の方法

複数検体を長鎖インサートのメイトペアライブラリを作成、次世代シーケンズ法でリードすることで、網羅的な均衡型構造多型・変異を同定する。具体的には、ゲノム DNA を PicoGreen で濃度調整後、ウオーターシェアーを 2 種類の長さ (3 kb、8 kb) で均等に切断、DNA Blunting Kit により切断片を平滑化する。3' 端が CAGCAG で一部にビオチン標識されたインターナルアダプターを T4 DNA Ligase でライゲーションした後、DNA を環状化する。この環状 DNA を EcoP15I で切断 (もしくは別の方法で) すると中央にアダプターが、両端にペアとなった 25-27 塩基のターゲット部位を含む DNA 断片が得られる。本申請書記入時点では、よりシーケンズ精度の高い Life Technologies/SOLiD™ 4 System が導入されておらず、Illumina/Solexa による方法の記載とするが、前者が導入された後は、前者によるプロトコルの最適化を行う。ビオチンタグにより DNA を濃縮し、アダプターを付加し、次世代シーケンズ法による塩基配列の決定を行う。ここで使用するアダプタ

ーは、検体識別のためのタグを付加し最低 12 検体のシーケンスを同時に実施する。

データ解析は、均衡型構造多型・変異検出専用プログラムで実施する。まず、タグ配列等を除去し、リピートマスクによりターゲット部位(1、2とする)のリピートをマスク、ターゲット部位1をリファレンスゲノムへマッピングさせる。マッピング可能な場合、その周囲配列(3 kb 断片では片側 5 kb、8 kb 断片では片側 10 kb)を優先しマッピングするが、マッピング不可の場合、更に外側に、最終的には全ゲノムに対してマッピングさせる。ターゲット部位の方向がリファレンスに一致かつその距離も閾値以下である場合(あらかじめ DNA 断片の長さを電気泳動で概算しておく)“構造変化なし”、方向がリファレンスに一致かつ閾値を超えて大きい場合“挿入”、方向がリファレンスに一致かつ閾値を超えて小さい場合“欠失”とする。方向がリファレンスと不一致の場合“逆位(切断点を含む)”、ターゲット部位の一方(2)が近傍にマップしないが同じ染色体である場合“染色体内転座もしくは逆位”(ターゲット部位の方向に依存)、ターゲット部位の一方(2)が別染色体にマップした場合“染色体外転座”と判断する。ただし、単一メイトペアのみでは、正確な判断は不可能であるため、得られた情報を検体毎に物理的位置に応じてスコア化して解析する。得られたスコアに応じてこれらの均衡型構造多型・変異をリスト化し以下の解析で使用する。

既に日本人検体で逆位が既知の領域(chr12:121.97Mb 領域等)を用いて切断点同定法の検討を行う。リファレンスゲノムで同方向 primer を 1-4 kb 毎に設計し切断点を含む PCR 法により増幅が可能であることを確かめ、更に、multiplex PCR 法での増幅が可能か検討する。電気泳動で得られた各バンドを Sanger 法により塩基配列を確認、順次添加する primer 数を増やし領域がオーバーラップする multiplex PCR 法の効率を検討する。

次に同一逆位領域を対象として、Loop-Mediated Isothermal Amplification (LAMP)法による逆位領域の増幅を試みる。通常の LAMP 法通り 3' 側に増幅する配列に相補的な配列を、5' 側に同一配列を含む primer を設計、このような primer を逆位領域及びその周辺領域に同方向に複数設計しておく。ここで、外側 primer は逆位切断点より遠位の primer で代用する。作成された primer を隣接 4 本で 1 セットとすることで(primer 数 - 1)セットの primer セットを作成、LAMP 法による増幅を試みる。増幅された amplicon に対し内部 primer を用いて Sanger 法でシーケンスを行い、逆位部分が増幅されているかを検討する。更に、切断点未知の領域ではどの primer セットが増幅されるか不明であるため、これを multiplex 化して増幅可能かを検討する。(Molecular inversion probe) MIP 法では、片側の相補配列を元の配列と同

一となるように設計、これと RCA 法による増幅を組み合わせ、LAMP 法と同じような手順で逆位切断点の増幅が可能であるかを検討する。本研究細目は分担研究者である宮川が担当する。

3) 均衡型構造多型・変異地図の作製(馬淵)

最近、DNA チップの解析に基づいて逆位を検出する手法が開発された(J Comput Biol, 17:517-, 2010)。これは、コンピューター上の予測であり一定の長さ(200Kb)以上でかつ集団内に高い頻度(20%)で存在する逆位多型のみという制限があり、その領域も少数であるがこの候補領域を対象に、PCR 法もしくは研究細目 2)による方法で切断点の同定を行い、日本人の均衡型構造多型・変異データを収集する。

#### 4. 研究成果

ヒトゲノムプロジェクト、HapMap プロジェクトにより、ヒトゲノムに存在する SNP が数多く同定され、また、ヒトゲノム構造多型コンソーシアムをはじめとした多くの国際共同研究により、CNV(コピー数多型)が同定されてきた。しかしながら、その多くはゲノム上の大まかな領域の検出にとどまり、CNV の存在の実験レベルの確認や具体的な切断点の情報に欠く等の問題がある。また、均衡型構造変化である逆位は、DNA チップによる検出が困難でありまたその詳細な地図は存在していないのが現状である。本研究では、特に逆位、染色体内・染色体外転座等の均衡型構造多型・変異が生じている部分の網羅的検出とその切断点の同定を目指して、その方法論を確立すること、及び日本人検体を用いてその均衡型構造多型・変異地図を作成することが本研究の目的である。このため、複数検体を長鎖インサートのメイトペアライブラリを作成、次世代シーケンス法でリードすることで、網羅的な均衡型構造多型・変異を同定する。具体的には、ゲノム DNA を PicoGreen で濃度調整後、ウォーターシェアーを 2 種類の長さ(3 kb、8 kb)で均等に切断、DNA Blunting Kit により切断片を平滑化、3' 端が CAGCAG で一部にビオチン標識されたインターナルアダプターを T4 DNA Ligase でライゲーションした後、DNA を環状化した。この環状 DNA を EcoP15I で切断すると中央にアダプターが、両端にペアとなった 25- 27 塩基のターゲット部位を含む DNA 断片が得られる。この後、ビオチンタグにより DNA を濃縮し、アダプターを付加し、次世代シーケンス法による塩基配列の決定を行った。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 0 件)

〔学会発表〕(計 0件)

〔図書〕(計 0件)

〔産業財産権〕  
出願状況(計 0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

馬淵 昭彦(MABUCHI AKIHIKO)  
東京大学・大学院医学系研究科・准教授  
研究者番号：80312312

### (2) 研究分担者

宮川 卓(MIYAGAWA TAKU)  
東京大学・大学院医学系研究科・助教  
研究者番号：20512263