

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 10 日現在

機関番号：63801

研究種目：基盤研究（B）

研究期間：2009～2011

課題番号：21310126

研究課題名（和文） 個人ゲノムの多様性に関する研究

研究課題名（英文） The Study on Human Genome diversity

研究代表者

豊田 敦（TOYODA ATSUSHI）

国立遺伝学研究所・生命情報研究センター・特任准教授

研究者番号：10267495

研究成果の概要（和文）：本研究では、遺伝的多様性に基づいて個人医療を進めるための基盤技術の開発を目的とし、次世代シーケンサー（Illumina）を用いて ①B リンパ球株化細胞から染色体特異的に分離した 21 番染色体と Y 染色体のゲノム解析と ②HapMap 計画で使用された日本人由来の細胞株の全ゲノム解析を実施した。その結果、日本人の配列解析から 300 万サイトを越える一塩基多型（SNVs）および 33 万サイトを越える挿入・欠失領域を検出するとともに公開されているアジア人の多型情報との比較を行った。今後、さらに多様性（特に構造多型）を高精度に同定するためには、シーケンス反応の最適化やメイトライブラリー作製法などの開発が必要である。また、SNVs の検出においても既存の解析ツールでは擬陽性が含まれて場合があり、プログラムの開発も重要な課題である。以上の課題を達成することにより、ゲノム情報に基づく医療体制の基盤が構築可能になると考えている。

研究成果の概要（英文）：Next generation sequencing offers lower cost and higher throughput than Sanger sequencing for human personal genome. To understand all genetic variants within personal genome, we carried out re-sequencing of two chromosomes (HSA21 and HSAY) separated from a human cell line and a sample from the HapMap JPT (Japanese) sets using the GAIIX sequencers. In a Japanese male genome, we identified over 3 million sequence variants (SNVs) at an average of 30-fold coverage. In addition, not only SNVs but also structural variants, such as indels and copy number variants, were detected.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	7,000,000	2,100,000	9,100,000
2010 年度	5,000,000	1,500,000	6,500,000
2011 年度	1,300,000	390,000	1,690,000
総計	13,300,000	3,990,000	17,290,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・基礎ゲノム科学

キーワード：個人ゲノム・多様性・次世代シーケンサー

## 1. 研究開始当初の背景

2004 年に解読が完了したヒトゲノム配列はテロメア領域をはじめクローンギャップやシーケンスギャップなどま

だ不完全な領域が存在し、現在も検証・更新作業が進められている。

これまでに標準配列を使用して比較的高頻度の多型が整備されたことによ

り、遺伝的要因と環境要因が複雑に関与している一般的疾患の原因となる DNA 変異を迅速に同定することが可能となってきている。しかし、まだ多くの一般的疾患において同定された多型の相対危険度が非常に低いことから、対立頻度は低いが発症に関与している DNA 変異（多型）の存在が示唆されている。そのため、現在においても個人ゲノムの再シーケンス解析（およびエキソーム解析）には大きな期待が寄せられている。

また、個人ゲノム解析が可能となった背景には、低コストかつ配列決定処理能力がこれまでとは桁違いである次世代シーケンサーと呼ばれる、サンガー法とは異なる原理による超並列配列決定装置が登場したことによる。さらに今後数年以内には、ヒトゲノムを 1000 ドル程度で解析する第三世代、第四世代のシーケンサーの実用化も視野に入っており、医療への応用を目指した個人ゲノム解析が飛躍的に進むことが予想される（米国では 1000 ドルゲノムプロジェクトが進められている）。研究開始当初において、すでにワトソン博士を含む数人程度の個人ゲノム配列が相次いで公開され、米国・英国・中国を中心に国際協力のもと「1000 人ゲノムプロジェクト」が開始された。このパイロット計画では世界中から 1000 人程度のゲノムを解読（～4 倍程度のカバーレッジ）し、遺伝的多様性を明らかにすることであった。その成果は、2011 年 10 月末に発表され、現在も解析規模を 2500 人に拡張してプロジェクトを進めている。また、個人間の違いがこれまで考えられていたよりも大きく、さらに個人間の多様性は一塩基多型だけではなく、構造多型（逆位、挿入、欠失、重複および遺伝子のコピー数の違い）などさまざまなゲノム構造変化が関与していることが明らかとなってきた。

## 2. 研究の目的

個人の全ゲノム配列情報すなわち遺伝的多様性に基づいて治療や予防を推進するためには、我が国でもその基盤技術を構築しておく必要がある。ゲノム構造の違いは、疾患ばかりではなく体質や薬物応答なども密接に関連しており、新たな予防や治療法の基盤情報として明らかにすることが必要不可欠である。また、同定された DNA 構造の違いが日本人の集団では見いだせない（関与していない）場合が多いため、独自に日本人の多型情報およびゲノム配列決定を進める必要がある（なお、国際コンソーシア

ムが決定した標準配列は、複数のリソース・ゲノムが用いられている）。

そこで本研究では、次世代シーケンサー（GAIIx: Illumina）を用いてシーケンス用の鋳型調製法の最適化やアリル間の違いの検出、マッピング・アセンブリプログラムの条件などを検討し、ヒトゲノムの多様性を探るためのさまざまな多型情報を整備することを目的とした。さらに個人の診断に用いるためには、簡便で極微量の DNA から解析することが重要である。また、1 細胞からゲノム配列を決定することが可能になれば、生殖細胞における卵や精子形成過程での複製エラーや突然変異、ゲノム構造変化（組換え）を塩基レベルで検出することが可能となる。そこで、極微量の DNA からの配列決定法の検討もあわせて行った。

## 3. 研究の方法

本研究では、ヒトにおける遺伝的多様性（個人差）を明らかにするための基盤技術開発を目的として、ヒト細胞株からセルソーターを用いて特定の染色体を分離し、次世代シーケンサー（GAIIx: Illumina）によりペアエンドシーケンスを行った。その際、極微量の DNA からの配列決定方法についても検討した。

また、決定した配列データをアセンブリすることにより得られたコンティグ配列や標準配列に配列データをマッピングした結果から一塩基多型および構造多型などを検出するための条件を検討した。検出された遺伝子領域の一塩基多型および遺伝子コピー数多型については、DNA チップの結果や dbSNP と比較するとともに多型情報をデータベース化した。

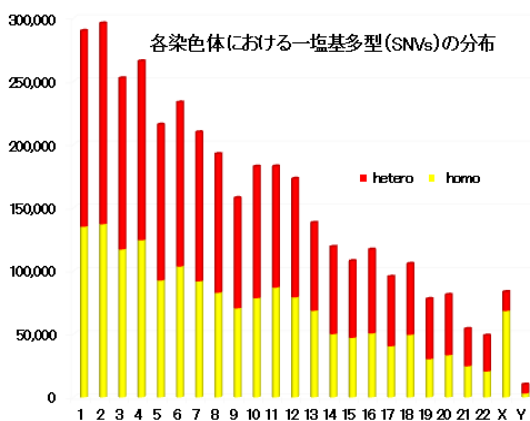
## 4. 研究成果

本研究では、B リンパ球株化細胞（GM130B）からセルソーターにより染色体特異的に単離したヒト 21 番染色体（約 350 万本の染色体）と Y 染色体由来（約 250 万本の染色体）のゲノム DNA を用いてペアエンドシーケンスを実施した。その過程において、ゲノム DNA の純度や必要量、コバリスを用いたゲノム DNA の断片化条件と DNA 測定、PCR 条件、qPCR による定量などの工程を最適化することにより、ペアエンドシーケンス用の鋳型 DNA ライブラリーの調製法を確立した。なお、最近では鋳型調製時において PCR を実施しない方法を使用しているが、必要なゲノム DNA 量が通常の 2 倍程度

(2-3ug)である。また、微量 DNA サンプルのシーケンスについては、全ゲノム増幅による条件検討を行ったが、バイアスやPCRによる高頻度な重複、キメラ配列などが認められ、現在もゲノムおよび鋳型調製法を検討している。

ランについては、高精度な配列の取得に向けて鋳型DNA量やクラスター数の最適条件を決定するとともに、配列評価用のパイプライン(評価項目:パスフィルター率・ユニークマップ率・PCRによる重複率・コンタミ率・リード上の精度値の分布・インサート長の分布・リード上のエラー率の分布など)を構築した。また、種々のプログラムによる多型の検出やアセンブリ結果を比較することにより解析パイプラインを整備した。各染色体特異的ゲノムDNAを使用したシーケンスでは、どちらも全リード(21番染色体のサンプルでは合計約209.1百万リード、Y染色体は約109.7百万リード)の約半分程度が目的とする染色体(Build37)にマップされ、残りのリードは他の染色体および株化する際に使用されたEBウイルス由来であった。21番染色体上にユニークにマップされたリードの重複度とカバーレッジは、それぞれ約87.3倍と約99.8%に対し、Y染色体では、約14.3倍と約55.0%であり、マルチヒットを許した場合の約半分程度(約35.8倍、約99.6%)であった。この結果は、Y染色体特有のゲノム構造(巨大な繰り返し配列の存在)が原因であり、解読長が短い配列によるゲノム解析が非常に困難であることを示している。なお、一塩基多型(SNVs)は、21番染色体上から約5.2万サイト(ホモとヘテロの合計)見出された。

また、次のサンプル(HapMap計画で使用された日本人由来の細胞株)のゲノム解析については、①次世代シーケンサーの処理能力の向上および低コスト化が進んだこと、②セルソーターでの染色体の分離に時間かかる(律速になる)こと、



③純度があまり良くないなどの理由から全ゲノムを対象に配列決定を試みた。

ヒト標準配列にマップした結果、全ゲノムの約97.5%をカバーしており、リードの重複度は平均29.3倍であった。つぎにSNVsを調べた結果、ホモとヘテロをあわせて約324万カ所が見いだされ、そのうちの約90%以上がSNPデータベース(db129)と一致していた。なお、これらの結果は、これまでの報告と大きく変わっていない。また、公開されている中国人(YH)や韓国人(SJK, AK1)の多型情報と比較した結果、韓国人と共通なSNVsが中国人よりも7-8%程度多かったが、さらに詳細な検討が必要である。

エクソン領域上には約26,000 SNVsが検出され、その約半分が非同義置換に関与していた。興味深いことにCDS領域上に終始コドンが存在する約90個の転写産物も見つかっており、現在検証を進めている。また、比較的小さな(~10塩基程度)挿入・欠失領域は、約33.3万サイト(ホモとヘテロの合計)存在していた。遺伝子コピー数多型および大きな挿入・欠失領域については、DNAチップの結果と比較することにより候補領域を抽出しており、qPCRなどで検証する予定である。

今後さらに次世代シーケンサーが苦手とするリピートや構造多型(特に遺伝子コピー数多型や大きな挿入など)の検出精度をあげるために、第三世代のシーケンサーなどの利用を検討していきたい。

## 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5件)

- ① Kim RN, Kim DS, Choi SH, Yoon BH, Kang A, Nam SH, Kim DW, Kim JJ, Ha JH, Toyoda A, Fujiyama A, Kim A, Kim MY, Park KH, Lee KS, Park HS. Genome analysis of the domestic dog (korean jindo) by massively parallel sequencing. *DNA Res.* 19(3):275-288, 2012 (査読有)
- ② Izutsu M, Zhou J, Sugiyama Y, Nishimura O, Aizu T, Toyoda A, Fujiyama A, Agata K, Fuse N. Genome features of "Dark-fly", a *Drosophila* line reared long-term in a dark environment. *PLoS One.* 7(3):e33288, 2012 (査読有)
- ③ Tsukahara S, Kawabe A, Kobayashi A, Ito T, Aizu T, Shin-i T, Toyoda A, Fujiyama A, Tarutani Y, Kakutani T. Centromere-targeted de novo

integrations of an LTR retrotransposon of *Arabidopsis lyrata*. *Genes Dev.* 26(7):705-713, 2012 (査読有)

- ④ Nishito Y, Osana Y, Hachiya T, Pependorf K, Toyoda A, Fujiyama A, Itaya M, Sakakibara Y. Whole genome assembly of a natto production strain *Bacillus subtilis natto* from very short read data. *BMC Genomics.* 11:243, 2010 (査読有)
- ⑤ 水口洋平・豊田敦, 1000 ゲノムプロジェクトの目標と展開, 医学のあゆみ, 236, 616-622, 2011 (査読無)

〔学会発表〕(計 1 件)

- ① 豊田敦, 次世代シーケンサーを用いた全ゲノム解析, 日本人類遺伝学会第 55 回大会, 平成 22 年 10 月 30 日, 大宮ソニックシティホール 大ホール

## 6. 研究組織

### (1) 研究代表者

豊田 敦 (TOYODA ATSUSHI)

国立遺伝学研究所・生命情報研究センター・特任准教授

研究者番号：10267495