

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 17 日現在

機関番号：63801

研究種目：新学術領域研究（研究領域提案型）

研究期間：2010～2014

課題番号：22129003

研究課題名（和文）次世代シーケンサーを用いたパーソナルゲノム解析技術の開発研究

研究課題名（英文）Study on Next-Generation Technologies for Personal Genome Analysis

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

交付決定額（研究期間全体）：（直接経費） 98,600,000円

研究成果の概要（和文）：本研究では、次世代型シーケンサー（Illumina）を利用してパーソナルゲノム上の多型を検出するための技術開発を行い、その技術を用いておもに家系情報のある脳神経疾患（アルツハイマー病、パーキンソン病、筋萎縮性側索硬化症、脊髄小脳変性症、統合失調症など）患者の全ゲノム配列決定を実施した。筋萎縮性側索硬化症については、新規原因遺伝子であるERBB4(ALS19)を同定した。また、繰り返し配列の異常伸長や遺伝子コピー数多型、大きな挿入・欠失などの構造多型を精度高く検出するために、ロングリード（PacBio）の鋳型調製法や解析技術の開発を実施した。

研究成果の概要（英文）：For the past ten years, next generation sequencing has advanced rapidly and is great potential means in the treatment and prevention of disease. Therefore, in this study, we developed sequencing technologies for personal genome analysis including brain diseases (Alzheimer disease, Parkinson disease, amyotrophic lateral sclerosis, spinocerebellar degeneration, and schizophrenia) and Japanese genetic variants were detected with the Illumina short reads. The novel mutation in ERBB4 gene was identified from a Japanese family affected by late-onset, autosomal-dominant amyotrophic lateral sclerosis using a whole-genome sequencing and parametric linkage analysis. Furthermore, to detect structural variants with high accuracy, such as copy number variation and large insertion/deletion, we have developed sample preparation and information analysis for the PacBio long reads.

研究分野：ゲノム科学

キーワード：パーソナルゲノム 次世代型シーケンサー

### 1. 研究開始当初の背景

疾患の発症には、一塩基多型や構造多型、エピジェネティックなどのさまざまなゲノム上の変化が関与している。これまでの研究では、ゲノム上に存在する頻度の高い SNPs 情報をもとに候補領域を特定し、その領域内にあるすべての遺伝子について疾患との関連を調べるゲノムワイド関連解析が行われている。このアプローチは「common disease common variants 仮説」に基づくものであるが、見出される疾患感受性遺伝子が病態への寄与が極めて小さいことが多い。このことは、疾患のリスクを高める頻度の低い多型の存在や遺伝子のコピー数変異などの構造多型が関与していると考えられる。また、本手法では大きな家系の存在が必須となるため、疾患関連遺伝子の同定に至らない場合が数多く存在している。そこで、家系を集積せずに特異的に存在するあるいは頻度が低い変異（多型）を検出・同定していくためには、塩基レベルでの違いを網羅的に抽出し候補を絞り込んでいく全ゲノム配列決定が有効である。また、これまで同定された DNA 構造の違いが日本人集団では見いだせない場合も多いため、独自に日本人パーソナルゲノム配列を決定する必要がある。

これまでは、コストやデータ生産量などの点でパーソナルゲノム解析は非現実的であったが、最近になって次世代型シーケンサーと呼ばれる革新的な超並列型の配列決定装置がつぎつぎと登場したことにより実現可能となっている。また、2008 年から開始された「1000 人ゲノムプロジェクト」では世界中から 1000 人程度のゲノムを解読（現在ではさらに解析数を増やしている）し、遺伝的多様性（集団において 1%程度で存在する多型）を明らかにすることを目標としている。しかしながら、配列データのカバレッジが低いいため、本研究が目的としているような稀な多型を見出すことは困難であることが予想され

る。また、国際コンソーシアムが決定した標準配列は多くのリソースを用いて構築されており、日本人由来のパーソナルゲノム配列はまだ決定されていない状況である。

### 2. 研究の目的

本領域研究では、疾患の原因遺伝子の解明および遺伝的多様性に基づくパーソナルゲノム医療の実現に向けて、(1) 次世代型シーケンサーによる高精度な配列決定を目指した技術開発、(2) 脳神経疾患患者のゲノム配列決定と多型の検出、(3) 日本人ゲノムの多型データベースの構築、(4) 発症との関連を解明することを目的としている。とくに本領域研究で見出されたゲノムの多様性情報（多型の頻度情報）は広く公開することにより、多くの疾患の解明につながるとともに治療や予防に関する研究が革新的に進むことが期待される。また、創薬開発の基盤情報としても必要不可欠であり、薬剤の効果や副作用といった研究開発において重要な研究資源となる。

本研究では、次世代型シーケンサーを用いたゲノム解析技術開発を担当するとともに、その技術を用いておもに家系情報のある脳神経疾患（アルツハイマー病、パーキンソン病、筋萎縮性側索硬化症、脊髄小脳変性症、統合失調症など）患者の全ゲノム配列を決定することにより疾患に関与する原因遺伝子の同定や発症機構の解明を目的とした。

### 3. 研究の方法

パーソナルゲノム医療を目指したゲノム解析が可能となった背景には、短時間に膨大な配列決定を可能にした次世代型のシーケンサーの登場があげられる。しかしながら、頻度は低いが高リスクの高い変異を見出すためには、ヒトゲノム上に存在する多型を高感度、高精度に取得することが求められ、配列の精度や配列解析プロトコールの整備が重要な課題である。また、次世代型シーケンサ

ーは技術の進展が早いため、その特性を見極めつつ、それぞれに最適な実験系を確立していく必要がある。そこで、ゲノム配列決定以外のさまざまなアプリケーション(メチル化解析など)にも対応できる大規模ゲノム解析システムを構築するとともに、ゲノム DNA の断片化、鋳型調製時におけるバイアスや反応効率の改良、鋳型 DNA の定量および極微量の DNA からの調製について検討した。

また最近、未解明な脳疾患においては、繰り返し配列の異常伸長や構造変異(遺伝子コピー数多型など)が原因の1つではないかと考えられており、ショートリード(HiSeq データ)だけでは検出が非常に困難な状況である。そのため、ロングリード(PacBio や Moleculo 技術)を用いたゲノム解析技術の開発をヒトや微生物ゲノムなどを用いて実施した。さらに、効率的なターゲット領域の濃縮やイルミナ技術によりエンドシーケンスが可能な fosmid ベクターの開発も合わせて検討を行った。これらの改善点を生かして脳疾患の検体について全ゲノム配列決定を実施した。

#### 4. 研究成果

HiSeq2000 (その後、HiSeq2500 にアップグレード)を導入し、高精度かつ高い処理能力をもつ配列決定システムを構築するとともにペアエンドシーケンス用のインサートサイズが均一でバイアスがかかりにくい(増幅しない)鋳型調製法を開発した。さらに、ゲノム構造多型を検出するためには種々のメイトペア情報の活用が有効であり、メイトペア間の距離が 3kb-15kb 程度までについては距離の分散が小さいライブラリ作製法を確立した。Fosmid エンド用のベクター作製についてはタイター効率が悪いいため、現在原因を究明中である。また、現行のメイトペアの作製法では多くの DNA を必要とするため改良を進めている。本研究で確立したペアエンドとメイトペアの作製法は、非モデル生物などの

新規ゲノム配列決定(de novo assembly)においても非常に有効であった。しかしながら、繰り返し配列の異常伸長や遺伝子コピー数多型、大きな挿入・欠失、重複領域などの検出(特にヘテロの場合)には、ショートリード(HiSeq データ)だけでは非常に困難な状況である。そこで、ロングリード(PacBio や Moleculo 技術)を用いた解析手法についてもヒトや微生物ゲノムを用いて検討することにより鋳型調製法や解析に必要なデータ量、検出精度の条件などを確立した。今後、ロングリードを用いて日本人ゲノムのハプロタイプ別のゲノム配列決定を実施することにより標準配列を構築していきたいと考えている(候補領域の検証には、最近市販された1分子 DNA マッピングシステムなどを使用する)。

本領域研究内の研究代表者から提供された脳疾患由来の検体については全ゲノム配列決定を完了させ、得られたデータは日本人ゲノムの多型データベース(頻度情報)を作成するために、本領域内の柱の一つであるゲノム情報解析研究グループに提供した。情報解析については、ベースコールを含むデータ処理を迅速に実行するために解析サーバーの増強および解析パイプラインを構築した。

#### 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2件)

Takahashi Y, Fukuda Y, Yoshimura J, Toyoda A, (以下 32 名), Fujiyama A, Morishita S, Goto J, Tsuji S. ERBB4 mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis type 19. Am J Hum Genet. 2013, 93(5), 900-905. 査読有  
doi: 10.1016/j.ajhg.2013.09.008.

Miyake K, Yang C, Minakuchi Y, Ohori K,

Soutome M, Hirasawa T, Kazuki Y, Adachi N, Suzuki S, Itoh M, Goto YI, Andoh T, Kurosawa H, Oshimura M, Sasaki M, Toyoda A, Kubota T. Comparison of Genomic and Epigenomic Expression in Monozygotic Twins Discordant for Rett Syndrome. PLoS One. 2013, 8(6), e66729. 査読有  
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0066729>

〔学会発表〕(計3件)

Takahashi Y, Fukuda Y, Yoshimura J, Toyoda A, Kurppa K, Moritoyo H, Belzil V, Elenius K, Rouleau G, Fujiyama A, Morishita S, Goto J, Tsuji S. Mutations in ERBB4 That Disrupt The NRG-ErbB4 Pathway Cause Autosomal Dominant Familial ALS Type 19. American Academy of Neurology, (April 26 - May 3, 2014, Philadelphia, PA)

Fujiyama A, Kasahara M, Toyoda A. Incorporation of Long-Read Technologies into Hybrid-de novo Genome Sequencing Strategy: An Assessment on PacBio and Moleculo Read. PLANT & ANIMAL GENOME ASIA 2014 (May 19-21, 2014, SINGAPORE)

高橋祐二、福田陽子、豊田敦、日笠幸一郎、吉村淳、Kari Kurppa、森豊浩代子、Veronique V. Belzil、石浦浩之、三井純、JaCALS、祖父江元、西澤正豊、野元正弘、Klaus Elenius、Guy A. Rouleau、藤山秋佐夫、森下真一、後藤順、辻省次。筋萎縮性側索硬化症の新規原因遺伝子 ERBB4(ALS19)の同定。第55回日本神経学会学術大会 2014年5月21日-24日 福岡国際会議場・福岡サンパレス・福岡国際センター

〔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0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
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

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

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：