

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 23 日現在

機関番号：12601

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：22129001

研究課題名（和文）パーソナルゲノム情報に基づく脳疾患メカニズムの解明

研究課題名（英文）elucidation of mechanisms of brain diseases based on personal genome information

研究代表者

辻 省次（Tsuji, Shoji）

東京大学・医学部附属病院・教授

研究者番号：70150612

交付決定額（研究期間全体）：（直接経費） 283,500,000円

研究成果の概要（和文）：次世代シーケンサーを導入し、シーケンス技術、ゲノムインフォマティクス基盤を整備し、日本人ゲノムの標準ゲノムの作成、variationに関するデータベースの整備を進め、本研究領域全体で有効活用した。日本人ゲノムの参照配列の整備については、Pacific Biosciencesのシーケンサーを用いて、8-10kbp程度のlong readのデータを集積し、HiSeqによるerror補正とあわせることで、構造多型を含めた高精度な日本人ゲノムの参照配列の構築を進めた。神経疾患において観察される非コード領域のリピート配列の異常伸長を検出する新たなアルゴリズムを開発し、領域内で研究に活用した。

研究成果の概要（英文）：We introduced next generation sequencers, and developed various tools and algorithms for genome informatics to analyze data generated by next generation sequencers. These facilities and genome informatics are shared by the researchers in this research project. To establish reference genome sequence of Japanese genome, we developed a new algorithm employing long reads obtained using the Pacific Biosciences sequencer along with error correction using short reads obtained using HiSeq sequencers. This algorithm facilitated detection of structural variations. To detect expanded short tandem repeats that are frequently associated with neurodegenerative diseases, a new algorithm to detect expanded short tandem repeats was developed, which was shared by the researchers in this research project.

研究分野：分子遺伝学，神経内科学

キーワード：ゲノム 個人ゲノム 脳疾患 インフォマティクス シーケンサー

### 1. 研究開始当初の背景

遺伝性脳疾患の病態解明においては、候補領域の十分な絞り込みができないために、これまでのゲノム解析技術では病因遺伝子の同定が著しく困難な疾患が少なからず残されている。孤発性疾患については、誰もが持っているようなゲノム多様性が疾患発症に關与するという考え (common disease-common variants 仮説) に立って 50-100 万個の SNPs を用いたゲノムワイド関連解析研究 (GWAS) が精力的に進められてきている。しかしながら、GWAS で見出される疾患感受性遺伝子は、odds 比が小さく、疾患の病態機序全体を理解するには至っていない。最近になり、疾患の病態機序に大きな役割を果たすゲノム多様性 (variants) は、アレル頻度の低いものであって、このような低頻度の variants は GWAS では検出し得ないということに気づかれ始めた。このような低頻度の variants の検出は GWAS では理論的に不可能であり、塩基配列レベルの解析が必須となる。脳疾患は、臨床病理学的に同一と考えられる疾患が、孤発性にも家族性にも觀察されることから、rare variants の探索を進める研究の対象として、最適であると考えられる。

次世代シーケンサーの技術開発が進み、最近になりようやく実用化された。次世代シーケンサーの技術については、数多くの課題が存在する。特に、本研究領域との関連では、解析精度が課題となっている。すなわち、ヒトゲノム上に存在する低頻度の variants を高精度に取得することが課題となっている。

インフォマティクスの分野は、わが国ではもっとも研究者層の薄い領域で、特に、ゲノムインフォマティクス、シーケンシング技術、生物学・医学研究の分野に横断的に精通したインフォマティシャンは極めて限られている。国際的にはこの領域には優れた研究者が多く、わが国が力を入れる必要のある領域である。ヒトゲノム多様性解析での課題は大きく分けて3つある。第1は、20兆塩基対前後の情報量を扱うアルゴリズムと並列DB処理技術の研究開発である。第2は、次世代シーケンサーの威力で rare variants の検出、逆位、転座などの大規模な構造的変化の検出が可能となったことである。検出の精度については十分な考察がされていない過渡期にある。第3は、多様性をいかにして機能的に分類するかという点である。variants は無数と言えるほど膨大にあり、そこから機能的に重要な variants を検出し機能ごとに分類するための統計学、データマイニング、システムズ生物学を組合せた新しい技術を開拓する必要がある。

### 2. 研究の目的

本研究の目的は、実用化されたばかりである次世代シーケンサーをコアとする最先端のゲノム解析技術研究、次世代シーケン

サーによって産出される膨大な情報に対する最先端のインフォマティクス研究、ヒトの代表的な脳疾患の高精度の全ゲノム配列解析 (パーソナルゲノム解析) に基づいて、疾患の発症に關与するゲノムの多様性 (common variants から rare variants までを含む) を明らかにし、発症機構を解明し、治療法開発の基盤構築の実現をめざすゲノム医学研究、という3つの最先端研究領域を融合することにより、全く新しい研究領域を創成し、将来のゲノム科学・医学研究を飛躍的に発展させていくことにある。

### 3. 研究の方法

本研究領域では、脳疾患の病態機序の根本的な解明を実現するためには、従来の、頻度の高い多型性 (common SNPs) に着目した GWAS では困難であって、疾患の発症に大きなインパクト (effect size が大) を持つ rare variants の探索が重要であると考え、その実現には、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研究、高度のゲノムインフォマティクス研究が必須であるという考え方に立ち、1. 次世代シーケンサーによるゲノム解析技術研究を進め、この成果に基づき、2. rare variants の探索に適した脳疾患に焦点を置いて大規模ゲノム配列解析を実施し、3. 高度のゲノムインフォマティクス解析研究の成果に基づき、疾患発症に關与するゲノム上の rare variants を同定し、疾患の病態機序を解明することをめざした。

3つの研究項目間で密接に連携し、研究を推進することを目指して、1年に2回、研究班会議、総括班会議を開催し、成果を共有し、必要に応じて頻回の情報交換を進めた。

### 4. 研究成果

研究班会議を2回、総括班会議を2回開催して、本研究領域の推進について討議を行った。研究領域内で、次世代シーケンサー (HiSeq2000 (Illumina), 5500XL (Life Technologies), Pacific Biosciences) 導入、解析に必要なゲノムインフォマティクスのパイプラインを整備し、領域内で有効に活用できる体制を整備し、本研究領域で有効利用することにより研究を進めた。

ゲノムインフォマティクスのパイプラインについては、exome 解析、全ゲノム配列解析に基づいて、ゲノム配列状の variation の検出に関して、いかに精度を高めるかという技術的な課題について検討を行った。また、日本人ゲノムの参照配列の整備、日本人ゲノムの variation database に作成のために全ゲノム配列解析、exome 配列解析のデータ収集を進めた。日本人ゲノムの参照配列の整備については、Pacific Biosciences のシーケンサーを用いて、8-10kbp 程度の long read のデータを集積し、HiSeq2000 による error 補正とあわせることで、構造多型を含めた高精度な日本人ゲノムの参照配列の構築を進

めた。このようにして得られた、日本人ゲノムの variation 情報を in-house database として整備し、ゲノム配列情報の解釈において極めて有用なリソースとして活用した。

神経疾患においては、非コード領域のリピート配列の異常伸長が疾患発症の原因となっていることが見出されてきており、次世代シーケンサーを用いた解析で得られる short read の中から、異常伸長したリピート配列を検出する新たなアルゴリズムを開発し、このアルゴリズムを用いて、領域内で広く異常伸長したリピートの探索を進めた。また、新生突然変異の検出に関しては、トリオ解析を用いた高精度の新生突然変異の解析方法を確立し、領域内でこの技術を共用し、研究に活用した。

#### 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 19 件)

- Mitsui J, Matsukawa T, Sasaki H, Yabe I, Matsushima M, D'rr A, Brice A, Takashima H, Kikuchi A, Aoki M, Ishiura H, Yasuda T, Date H, Ahsan B, Iwata A, Goto J, Ichikawa Y, Nakahara Y, Momose Y, Takahashi Y, Hara K, Kakita A, Yamada M, Takahashi H, Onodera O, Nishizawa M, Watanabe H, Ito M, Sobue G, Ishikawa K, Mizusawa H, Kanai K, Hattori T, Kuwabara S, Arai K, Koyano S, Kuroiwa Y, Hasegawa K, Yuasa T, Yasui K, Nakashima K, Ito H, Izumi Y, Kaji R, Kato T, Kusunoki S, Osaki Y, Horiuchi M, Kondo T, Murayama S, Hattori N, Yamamoto M, Murata M, Satake W, Toda T, Filla A, Klockgether T, Wllner U, Nicholson G, Gilman S, Tanner CM, Kukull WA, Stern MB, Lee VM-Y, Trojanowski JQ, Masliah E, Low PA, Sandroni P, Ozelius LJ, Foroud T, and Tsuji S. Variants associated with Gaucher disease in multiple system atrophy. *Ann Clin Transl Neurol* 2: 417-426, 2015. DOI: 10.1002/acn3.185. 査読有
- Ishiura H, Takahashi Y, Hayashi T, Saito K, Furuya H, Watanabe M, Murata M, Suzuki M, Sugiura A, Sawai S, Shibuya K, Ueda N, Ichikawa Y, Kanazawa I and Tsuji S. Molecular epidemiology and clinical spectrum of hereditary spastic paraplegia in the Japanese population based on comprehensive mutational analyses. *J Hum Genet* 59:163-72, 2014. DOI: 10.1038/jhg.2013.139. 査読有
- Tsuji S. Neurogenomics view of neurological diseases. *Arch Neurol* 70:689-94, 2013 PubMed PMID: 23571861 doi 10.1001/jamaneuro1.2013.734.
- Isojima T, Doi K, Mitsui J, Oda Y, Tokuhiro E, Yasoda A, Yorifuji T, Horikawa R, Yoshimura J, Ishiura H, Morishita S, Tsuji S, and Kitanaka S. A recurrent de novo FAM111A mutation causes Kenny-Caffey syndrome type 2. *J Bone Mineral Res* 29:992-8, 2014. doi 10.1002/jbmr.2091.
- Ishiura H, Sako W, Yoshida M, Kawarai T, Tanabeb O, Goto J, Takahashi Y, Date H, Mitsui J, Ahsan B, Ichikawa Y, Iwata A, Yoshino H, Izumi Y, Fujita K, Maeda K, Goto S, Koizumi H, Morigaki R, Ikemura M, Yamauchi N, Murayama S, Nicholson GA, Ito H, Sobue G, Nakagawa N, Kaji R and Tsuji S. The TRK-Fused Gene Is Mutated in Hereditary Motor and Sensory Neuropathy with Proximal Dominant Involvement. *Am J Hum Genet* 91: 320-329, 2012 An Y,
- Matsukawa T, Asheuer M, Takahashi Y, Goto J, Suzuki Y, Shimozaawa N, Takano H, Onodera O, Nishizawa M, Aubourg P and Tsuji S. Identification of novel SNPs of ABCD1, ABCD2, ABCD3 and ABCD4 genes in patients with X-linked adrenoleukodystrophy (ALD) based on comprehensive resequencing and association studies with ALD phenotypes. *Neurogenet* 12: 41-50, 2011
- Ishiura H, Fukuda Y, Mitsui J, Nakahara Y, Ahsan B, Takahashi Y, Ichikawa Y, Goto J, Sakai T and Tsuji S. Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1. *Neurogenet* 12: 117-121, 2011
- Matsukawa T, Wang X, Liu R, Onuki Y, Kubota A, Hida A, Kowa H, Fukuda Y, Ishiura H, Mitsui J, Takahashi Y, Aoki S, Takizawa T, Shimizu J, Goto J, Proud CG and Tsuji S. Adult-onset leukoencephalopathies with vanishing white matter with novel missense mutations in EIF2B2, EIF2B3, and EIF2B5. *Neurogenet* 12: 259-261, 2011
- Toyoda A, Zhao C, Fujiyama A, Agata K. A colony multiplex quantitative PCR-Based 3S3DBC method and variations of it for screening DNA libraries. *PLoS One* 10(2): e0116997, 2015. 査読有
- 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* 8(6): e66729, 2013. 査読有
- Ichikawa K and Morishita S. A linear time algorithm for detecting long genomic regions enriched with a specific combination of epigenetic states. *BMC Genomics* 16 (Suppl 2): S8, 2015. 査読有
- Doi K, Monjo T, Hoang PH, Yoshimura J, Yurino H, Mitsui J, Ishiura H, Takahashi Y, Ichikawa Y, Goto J, Tsuji S, Morishita S. Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. *Bioinformatics*. 30(6): 815-22, 2014. 査読有
- Qu W, Hashimoto S, Shimada A,

- Nakatani Y, Ichikawa K, Saito TL, Ogoshi K, Matsushima K, Suzuki Y, Sugano S, Takeda H and Morishita S. Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns. **Genome Res** 22: 1419-1425, 2012
14. Ogoshi K, Hashimoto S, Nakatani Y, Qu W, Oshima K, Tokunaga K, Sugano S, Hattori M, Morishita S, and Matsushima K. Genome-wide profiling of DNA methylation in human cancer cells. **Genomics** 98: 280-287, 2011

〔産業財産権〕

該当なし

#### 6 . 研究組織

研究代表者 辻 省次 (Shoji Tsuji)

東京大学医学部附属病院教授

研究者番号：70150612

分担研究者 豊田 敦 (Atsushi Toyoda)

国立遺伝学研究所特任准教授

研究者番号：10267495

分担研究者 森下真一 (Shinichi Morishita)

東京大学大学院新領域創成科学研究科教授

研究者番号：90292854