

201442080A

厚生労働科学研究委託費

難治性疾患実用化研究事業

神経疾患の集中的な遺伝子解析及び
原因究明に関する研究

平成26年度 委託業務報告書

業務主任者 辻 省 次

平成27年(2015)3月

本報告書は、厚生労働省の厚生労働科学研究委託事業による委託業務として、国立大学法人東京大学医学部附属病院が実施した平成26年度「神経疾患の集中的な遺伝子解析及び原因究明に関する研究」の成果を取りまとめたものです。

目 次

| | |
|---|----|
| I. 委託業務成果報告（総括） | 1 |
| 神経疾患の集中的な遺伝子解析及び原因究明に関する研究 辻 省次 | |
| II. 委託業務成果報告（業務項目） | |
| 1. 神経疾患の集中的な遺伝子解析及び原因究明に関する研究 ゲノムインフォマティクス研究 森下真一 | 8 |
| 2. 神経疾患の集中的な遺伝子解析及び原因究明に関する研究 遺伝性神経疾患の病因遺伝子の解明 石浦浩之 | 11 |
| 3. 孤発性神経疾患の病因遺伝子の解明 三井 純 | 14 |
| 4. 神経疾患の病態機序の解明 辻 省次 | 16 |
| 5. 診断を目的とした網羅的遺伝子解析システムの構築と遺伝子診断の提供 後藤 順 | 19 |
| III. 学会等発表実績 | 22 |
| IV. 研究成果の刊行物・別刷 | 25 |

厚生労働科学研究委託費(難治性疾患実用化研究事業)

委託業務成果報告(総括)

神経疾患の集中的な遺伝子解析及び原因究明に関する研究

業務主任者 辻 省次

東京大学医学部附属病院神経内科教授

研究要旨

次世代シーケンサーを用いたゲノム配列解析で課題となる、インフォマティクス解析について、エクソーム解析の検出感度を高める解析方法整備を行った。また、1分子シーケンシングを用いた高精度のゲノム解析方法を確立した。遺伝性神経疾患、孤発性神経疾患に関連する遺伝子の解析を進め、多系統萎縮症の発症に関わる新たな遺伝子として、*GBA* を見出した。次世代シーケンサーを診断確定に応用するために必要な、品質管理や精度管理の検討を行った。

A. 研究目的

神経疾患(難病)を対象疾患群として、次世代シーケンサーを駆使したゲノム配列解析拠点の整備、次世代シーケンサーから得られる膨大な規模のゲノム配列情報を処理するゲノムインフォマティクス拠点の整備を基盤として、発症機構が未解明の数多くの神経疾患の発症に関与する遺伝子を網羅的に解明し、その分子病態機序の解明、病態抑止型治療法開発を実現することを目的としている。あわせて、診断未確定の神経疾患を対象として、一般研究班、学会と連携して、高精度の遺伝子診断システムにより診断を確定することを目的とする。

B. 研究方法

ゲノムインフォマティクス研究

エクソーム関連解析において、低頻度アレルに着目した関連解析では、多重検定による

有意水準を満たすためには、膨大な数の検体を要することから、その遺伝統計学的な解析についての検討を行った。神経疾患にしばしば見いだされる DNA 変異の1つに短鎖繰返配列(Short Tandem Repeat) がある。このような配列の伸長変異を、全ゲノム配列解析から効率よく検出出来るアルゴリズムを開発した。ゲノム解析を効率よく進めるために、計算機設備の増強を行った。

遺伝性神経疾患の病因遺伝子の解明

主に家族性筋萎縮性側索硬化症、遺伝性痙性対麻痺、家族性多系統萎縮症家系に関して家系収集を進め、ハイスループット連鎖解析システムにあわせ、エクソーム解析・全ゲノム配列解析を施行し、原因変異同定を目指した。現在まで蓄積されてきた対照者のエクソーム解析・全ゲノム配列解析データにあわせ、今年度新規に解析された対照例を合わせた、ま

た、変異の minor allele frequency については、各種公開データベースを用いた。

孤発性神経疾患の病因遺伝子の解明

国際多施設共同研究で収集された孤発性多系統萎縮症患者 969 例(日本 574 例, 欧州 223 例, 北米 172 例), 健常対照者 1509 例(日本 900 例, 欧州 315 例, 北米 294 例)に対して, エクソーム解析を含む網羅的なゲノム解析を行った。日本人孤発性筋萎縮性側索硬化症 319 例に対して, エクソーム解析を行い, 発症原因となっている変異を探索した。

神経疾患の病態機序の解明

網羅的な RNA 配列解析を行うために, HiSeq2500 を用いて, RNA-seq を行った。サンプルを 101 塩基ペアエンド法で塩基配列解析を行い, 得られた短鎖長配列については, TopHat を用いて, 参照配列 hg19 と参照トランスクリプトーム配列(gencode v14)を用いて alignment を行った。各遺伝子の発現量, アイソフォーム解析については Cufflinks を用いて行った。

診断を目的とした網羅的遺伝子解析システムの構築と遺伝子診断の提供

一連の作業工程を見直し, 作業の標準化と各工程における質の確認を行った。遺伝子診断の際に病原性変異の候補を検出し, 絞り込むための情報解析パイプラインを整備した。シーケンス・エラーについて検討するため, 435 例のエクソーム・データの特定の遺伝子群における変異の中から, dbSNP に登録されていない変異のみを抽出し, 全変異についてその有無をサンガーシーケンスで確認した。一般研究班, 学会との連携を呼びかけ, 遺伝子診断を主な目的とするエクソーム解析サービスを提供した。

(倫理面への配慮)

ヒトゲノム・遺伝子解析研究に関する倫理指針に基づいて研究を行った。

C. 研究結果

ゲノムインフォマティクス研究

約 6 万人のエクソーム・データを解析したデータベース Exome Aggregation Consortium (ExAC) を MIT Broad Institute が 2014 年 10 月に公開した。このデータベースに含まれるデータを活用することにより, アレル頻度 0.5% 程度の疾患関連 rare variant を比較的少数の罹患者約 500 名からでも検出可能であることを確認し, この分析をするための情報解析システムを研究開発した。

1 分子リアルタイムシーケンサーを用いると数千塩基長の DNA 断片を解読できるものの, エラー率が 15% 程度と高く, エラー補正が不可欠である。そこで被覆率が 20 倍程度でも短鎖繰返配列を検出できるように, Illumina 社のシーケンサーで 100 塩基程度の短い配列を大量に読むことでエラー補正する技術を研究・開発した。

必要十分な計算機設備を増強した。

遺伝性神経疾患の病因遺伝子の解明

家族性筋萎縮性側索硬化症としては, 総計 68 家系となり, うち SOD1 27 例, FUS 7 例, TARDBP 2 例と, 本邦においてはこの 3 つの遺伝子の変異が比較的多く認められることが判明した。その他, ERBB4, C9ORF72 に変異を 1 例ずつ認めた。

遺伝性痙性対麻痺については, 家族歴より常染色体劣性遺伝が疑われた症例と, 孤発例でも複合型である症例で変異が認められていない 144 症例に関して, エクソーム解析を行った。その結果, 58 例(約 40%)については既

知

の遺伝子に変異が認められた。変異が同定されなかった症例に関して、複数家系に共通するような候補遺伝子を見出しており、今後、この遺伝子について、validationを進める。

非常に稀な家族性多系統萎縮症について、既に見出している *COQ2* 遺伝子変異を有する家系を除く家系について、収集・解析を続けている。

孤発性神経疾患の病因遺伝子の解明

日本、欧州、北米の3シリーズにおける患者・対照群の *GBA* 遺伝子機能障害性変異のキャリア頻度についてメタ解析を行ったところ、オッズ比 2.44 (95%信頼区間 1.14-5.21), p 値 0.029 にて、*GBA* 遺伝子機能障害性変異は多系統萎縮症患者と関連した。サブタイプ別では、小脳症状優位型 (MSA-C) において強い関連 (p 値 0.0073) が得られた。

日本人孤発性筋萎縮性側索硬化症 319 例については、9 例 (3.1%) に遺伝性筋萎縮性側索硬化症 (常染色体優性遺伝) の病原性変異を認めた。

神経疾患の病態機序の解明

筋萎縮性側索硬化症患者における RNA processing 異常を検出するために、脊髄組織を用いた RNA-seq を検討しているが、組織量が限られることから、まずは大脳における RNA 代謝異常を検討することを目的に、大脳 (n=3)、小脳 (n=2) を用いて RNA-seq を行った。大脳のサンプルにおいて、27 個の有意に発現変動が認められる遺伝子を見出した。これらの変動遺伝子を確かめるために、さらに症例数を増やして RNA-seq を行うとともに、今後脊髄の RNA-seq を行うことを検討していく。

診断を目的とした網羅的遺伝子解析システムの構築と遺伝子診断の提供

病原性変異の候補を絞り込むためには、遺伝形式の仮定が前提になる。 *de novo* 変異を仮定した場合には、両親や非罹患同胞には存在せず、罹患者だけが有するヘテロ接合性変異を抽出するアルゴリズムを開発し、診断に応用した。絞り込みにおいては変異の頻度情報が最も重要な条件になるが、インハウスにおける対照健常者 600 例の頻度情報、および ExAC の頻度情報を参照し、絞り込みに活用できるようなパイプラインを整備した。既知の病原性変異、既知の遺伝子・表現型関連について、Human Gene Mutation Database (HGMD), OMIM に基づくデータを参照可能なパイプラインを整備した。従来の変異検出パイプラインでは、塩基置換変異、数塩基の欠失・重複変異までしか検出できなかったが、新たに Conifer を用いてエクソームのデータを解析して、大きな欠失・重複変異を検出するためのパイプラインを整備した。466 個の dbSNP に登録のない変異について、サンガーシーケンスで確認したところ 74.9% が真陽性であり、25.1% は偽陽性であった。変異検出パイプラインとして用いている samtools による Variant quality score による層別化では、score 50 以上で陽性的中率が 97.3%、score 20-50 で 52.8%、score 20 未満で 9.9% であった。

遺伝性ニューロパチー、HTLV-I 関連脊髄症などの難治性疾患研究班を担当する鹿児島大学神経内科と連携して、これまでに計 693 例の検体のエクソーム解析を実施した。

遺伝性痙性対麻痺の全国多施設共同研究体制である Japan Spastic Paraplegia Research Consortium (JASPAC) と連携して、これまでに計 291 例の検体のエクソーム解析を実施した。

運動失調症の病態解明と治療法開発に関

する研究班, 神経変性疾患に関する調査研究班などと連携して, エクソーム解析の提供を開始した。

D. 考察

インフォマティクス研究の推進により, エクソーム関連解析において, 低頻度アレルの検出力を確保することが課題であったが, 約6万人のエクソーム・データを解析したデータベース ExAC の中から, アジア人のデータを抽出し, 検出力を格段に高めることができた。この解析方法を応用することにより, 多くの孤発性神経疾患の疾患感受性遺伝子の探索が進展することが期待される。

1分子シーケンシングによる解析方法は, エラー率 15%と高いことが課題であったが, Illumina 社のシーケンサーで 100 塩基程度の短い配列で補正することにより, 高精度の配列解析が可能になった。このことにより, 構造変異や, 神経疾患に多く観察される短鎖繰返配列の伸長変異の検出に応用できるものと期待される。

遺伝性神経疾患の病因遺伝子の解明については, 解析対象となる家系のサイズが小さい, 家系数が少ないことが課題であるが, インハウスのデータベースの充実, 公的データベースを活用することにより, 新規変異の解釈の精度を高めることができた。このパイプラインにより, 候補遺伝子が見出されており, 今後, validation を進めていく。

孤発性神経疾患の病因遺伝子の解明については, 多系統萎縮症, 筋萎縮性側索硬化症を主な対象として研究を進めている。多系統萎縮症の発症に関連する遺伝子として, 新たに, *GBA* 遺伝子を見出すことができた。今後, ExAC のデータベースを活用することにより,

検出感度を高めた関連解析を進めていく。

神経疾患の病態機序の解明については, RNA-seq の解析方法を確立し, 大脳, 小脳を用いた解析を開始した。今後, 筋萎縮性側索硬化症の脊髄など, 病変部位の解析を進めていく。

診断を目的とした網羅的遺伝子解析システムの構築と遺伝子診断の提供については, 病原性変異の候補を効率よく絞り込むパイプラインを確立した。次世代シーケンサーは, 一定のエラーを伴うことが知られているが, 実際に, 466 個の dbSNP に登録のない変異について, サンガーシーケンスで確認したところ 74.9%が真陽性であり, 25.1%は偽陽性であった。このデータを用いて, ROC curve を作成し, 感度, 特異度を高める条件を検討し, 必要に応じて, variants call に反映するようにし, 変異の検出精度を高めることができた。今後, 診断確定を目的とした variants call に応用していく。

E. 結論

本研究では, インフォマティクス研究により, エクソーム関連解析の検出感度を高める方法, 1分子シーケンシングを用いた高精度の解析方法を確立できた。これらの解析方法を神経疾患のゲノム解析に応用していく。

遺伝性神経疾患の病因遺伝子の解明, 孤発性神経変性疾患の発症に関連する遺伝子の解明について, 感度の高い解析方法, パイプラインを構築でき, 研究を推進していく。孤発性多系統萎縮症について, 新たな疾患感受性遺伝子として, *GBA* を同定した。

RNA-seq による, 発現プロファイル解析を開始し, 病態機序研究に応用していく。

次世代シーケンサーによる解析を診断目的

とした解析に応用するための各種パイプラインを整備した。また、変異の病的意義の解釈に有用なインハウスデータベースを整備し、精度の高い遺伝子検査として提供していく。

F. 健康危険情報

該当せず。

G. 研究発表

1. 論文発表

- Mitsui J, et al. Variants associated with Gaucher disease in multiple system atrophy. *Ann Clin Transl Neurol.* 2015 (doi: 10.1002/acn3.185)
- Shimazaki H, Honda J, Naoi T, Namekawa M, Nakano I, Yazaki M, Nakamura K, Yoshida K, Ikeda S, Ishiura H, Fukuda Y, Takahashi Y, Goto J, Tsuji S, Takiyama Y. Autosomal-recessive complicated spastic paraplegia with a novel lysosomal trafficking regulator gene mutation. *J Neurol Neurosurg Psychiatry.* 2014; 85: 1024-8.
- Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, Forlani S, Rakowicz M, Schöls L, Mariotti C, van de Warrenburg BP, Orsi L, Giunti P, Filla A, Szymanski S, Klockgether T, Berciano J, Pandolfo M, Boesch S, Melegh B, Timmann D, Mandich P, Camuzat A; Clinical Research Consortium for Spinocerebellar Ataxia (CRC-SCA); EUROSCA network, Goto J, Ashizawa T, Cazeneuve C, Tsuji S, Pulst SM, Brusco A, Riess O, Brice A, Stevanin G. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain.* 2014; 137: 2444-55.
- Doi H, Ushiyama M, Baba T, Tani K,
- Shiina M, Ogata K, Miyatake S, Fukuda-Yuzawa Y, Tsuji S, Nakashima M, Tsurusaki Y, Miyake N, Saito H, Ikeda S, Tanaka F, Matsumoto N, Yoshida K. Late-onset spastic ataxia phenotype in a patient with a homozygous DDHD2 mutation. *Sci Rep.* 2014; 4: 7132.
- Kanbayashi T, Saito F, Matsukawa T, Oba H, Hokkoku K, Hatanaka Y, Tsuji S, Sonoo M. Adult-onset vanishing white matter disease with novel missense mutations in a subunit of translational regulator, EIF2B4. *Clin. Genet.* (doi: 10.1111/cge.12554.)
- Koh K, Kobayashi F, Miwa M, Shindo K, Isozaki E, Ishiura H, Tsuji S, Takiyama Y. Novel mutations in the PNPLA6 gene in Boucher-Neuhäuser syndrome. *J. Hum. Genet.* (doi: 10.1038/jhg.2015.3.)
- Saigoh K, Mitsui J, Hirano M, Shioyama M, Samukawa M, Ichikawa Y, Goto J, Tsuji S, Kusunoki S. The first Japanese familial case of spinocerebellar ataxia 23 with a novel mutation in the PDYN gene. *Parkinsonism Relat Disord.* 2015; 21: 332-4.
- 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. Rapid detection of expanded short tandem repeats in personal

genomics using hybrid sequencing. *Bioinformatics*. 2014; 30: 815-22.

- Kazuki Ichikawa, Shinichi Morishita. A simple but powerful heuristic method for accelerating k-means clustering of large-scale data in life science. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2014; 11: 681-92
 - Kishin Koh, Hiroyuki Ishiura, Michiaki Miwa, Jun Yoshimura, Jun Mitsui, Jun Goto, Shinichi Morishita, Shoji Tsuji, Yoshihisa Takiyama. Exome sequencing shows a novel de novo mutation in ATL1. *Neurology and Clinical Neuroscience* 2014; 2: 1-4
2. 学会発表
- Morishita S, et al. Completing CpG methylation statuses in human and vertebrate genomes by integrating SMRT sequencing kinetic data. The 64th Annual Meeting of the American Society of Human Genetics
 - Koike A, et al. Update and expansion of Human Variation Database in Japanese Database Integration Program. The 64th Annual Meeting of the American Society of Human Genetics
 - Satake W, et al. Exome Association Study and 2nd SNP-GWAS of Japanese Parkinson's disease. The 64th Annual Meeting of the American Society of Human Genetics
 - Ishiura H, et al. Exome analysis of 116 patients supposed to be autosomal recessive hereditary spastic paraplegia

established molecular diagnoses of 49 patients with broad genetic

- heterogeneities. The 64th Annual Meeting of the American Society of Human Genetics
- 田中 真生ら. 次世代シーケンサーを用いたミトコンドリア遺伝子変異の高感度定量的解析. 第 55 回日本神経学会
- 松川 敬志ら. 思春期/成人大脳型副腎白質ジストロフィーに対する造血幹細胞移植の臨床的効果の検討. 第 55 回日本神経学会
- 三井 純. 多系統萎縮症の疾患関連遺伝子の探索. 第 55 回日本神経学会
- 石浦 浩之. 常染色体劣性遺伝が疑われた遺伝性痙性対麻痺 104 例の遺伝子解析. 第 55 回日本神経学会
- 野妻 智嗣ら. エクソーム解析を用いた HAM 疾患感受性遺伝子の探索. 第 59 回日本人類遺伝学会
- 田邊 肇ら. Charcot-Marie-Tooth 病におけるミエリン関連蛋白を中心とした新規遺伝子探索の試み. 第 59 回日本人類遺伝学会
- 石原 聡ら. ニューロフィラメントに関連した Charcot-Marie-Tooth 病の新規原因遺伝子同定の試み. 第 59 回日本人類遺伝学会
- 成瀬 紘也ら. Exome 解析を用いた家族性 ALS の分子疫学の解明. 第 59 回日本人類遺伝学会
- 高柳 俊作ら. 家族性腫瘍疾患 VHL 病に対するゲノム医療の展開. 第 59 回日本人類遺伝学会
- 佐藤 奈穂子ら. 46,XY gonadal dysgenesis を伴う minifascicular

neuropathy における新規遺伝子変異の
同定. 第 59 回日本人類遺伝学会

- 高 紀信ら. Boucher-Neuhauser 症候群
の新規 PNPLA6 遺伝子変異. 第 59 回
日本人類遺伝学会
- 大崎 雅央ら. 常染色体劣性遺伝形式を
示す spastic ataxia の病因遺伝子の探
索. 第 59 回日本人類遺伝学会

H. 知的財産権の出願・登録状況(予定
を含む)

1. 特許取得 該当せず.
2. 実用新案登録 該当せず.
3. その他 該当せず.

厚生労働科学研究委託費(難治性疾患実用化研究事業)

委託業務成果報告(業務項目)

神経疾患の集中的な遺伝子解析及び原因究明に関する研究
(ゲノムインフォマティクス研究)

担当責任者 森下 真一

東京大学大学院 新領域創成科学研究科 情報生命科学専攻 教授

研究要旨

神経疾患に固有の DNA 変異を検出するために、神経疾患のゲノム情報解析の基盤を整える。具体的には以下の3課題に取り組んだ。1. 低頻度の疾患関連変異を発見するために、約6万人のエキソームデータベース(ExAC)を取り入れて、日本人の神経疾患罹患者から得られたエキソームデータとの間で全ゲノム相関解析(Genome-Wide Association Study) を実施するソフトウェアを研究開発。2. 神経疾患に多い短鎖繰返配列(Short Tandem Repeat) の網羅的データベースを構築するために、従来型の Illumina シーケンサーと1分子リアルタイムシーケンサー(Single Molecule RealTime Sequencer)を組合せて、短鎖繰返配列を同定できるアルゴリズムを研究開発。3. 並列計算機システムおよびデータ格納用サーバーを整備。

A. 研究目的

神経疾患に固有の DNA 変異を検出するために、コントロール DNA との比較対照研究を実施する。さらに神経疾患に頻繁にみられる大規模な構造変異の検出に取り組む。この目標のために必要な情報解析設備を整備する。

B. 研究方法

1. 比較対照研究により統計的に有意な DNA 変異を検出する際に注意しなければならないのは、候補となる DNA 変異数が非常に多くなることである。多重検定による有意水準の補正(Bonferroni 補正等)が不可欠となる。たとえば 50 万個の変異

候補を想定して有意水準 5%で帰無仮説を検定するのであれば、確率が 10^{-7} 以下であるか否かで検定する必要がある。この考え方は、全ゲノム相関解析が普及した過去 10 年間で 1 つの基準となったが、全ゲノムおよびエキソーム解析においても踏襲すべき基準である。本研究課題でも、この理論に沿って考察を進めた。この厳しい検定基準を満たす低頻度の疾患関連変異を見出すためには、罹患者およびコントロールデータを多数用意する必要がある。どの程度の数のデータであれば必要十分であるかを理論的に考察することも重要になる。

2. 神経疾患にしばしば見いだされる DNA 変異の1つに短鎖繰返配列(Short Tandem Repeat) がある。これは 2-10 塩基を単位とする配列が繰返しコピーされて伸長した配列である。正常人に比べて神経疾患罹患者において著しく長くなる短鎖繰返配列が、罹患者系の連鎖解析から見出されている。しかしヒトゲノム全体で短鎖繰返配列がどのように分布しているかの網羅的なデータは未整備である。その理由は、長さが 1000 塩基を越える繰返し配列が多く、従来型の 100 塩基程度の短い配列しか読めない DNA シーケンサーを使って検出することは殆ど不可能だからである。我々は1分子リアルタイムシーケンサーを用いて数千塩基長の DNA 断片を解読することにより、この問題にアプローチしている。
3. 1人のヒトゲノム情報を解析するために必要な DNA データ量(生データおよび解析データ)は、全ゲノムの場合は 400GB、全エクソームの場合はその約 1/10 である。また1人分の全ゲノムデータを解析するためにかかる計算時間は2-3日程度(18CPU コアの並列計算機で実行)である。新しいデータを恒常的に収集し分析するためには、計算機設備を更新する必要がある。

(倫理面への配慮)

ヒトゲノム・遺伝子解析研究に関する倫理指針に基づいて研究を行った。

C. 研究結果

1. 約 6 万人のエクソームデータを解析したデータベース ExAC を MIT Broad

Institute が 2014 年 10 月に公開した。欧米人、アフリカ人に加えて南アジア人、台湾人のデータも多数含まれた多様な民族の総合的なエクソームデータベースである。アレル頻度 0.5%程度の疾患関連 rare variant (低頻度の DNA 変異)を、比較的少数の罹患者約 500 名からでも検出可能であることが、統計学的考察から分かった。そこでこの分析をするための情報解析システムを研究開発した。

2. 1分子リアルタイムシーケンサーを用いると数千塩基長の DNA 断片を解読できるものの、エラー率が 15%程度と高く、エラー補正が不可欠である。幸いエラーがランダムに入るため、被覆率 50 倍以上で DNA、断片を解読すると、エラーの補正が可能である。しかし従来の Illumina 社のシーケンサーに比べると消耗品コストが 1 桁高くなるため、現実には実施が困難である。そこで被覆率が 20 倍程度でも短鎖繰返配列を検出できるように、Illumina 社のシーケンサーで 100 塩基程度の短い配列を大量に読むことでエラー補正する技術を研究開発した。さらにコストを抑えるために、DNA 全体を1分子リアルタイムシーケンサーで解読するのを避け、Illumina 社のシーケンサーで候補領域を絞り込んでから、1分子リアルタイムシーケンサーで解読するアルゴリズムを研究開発した。
3. 必要十分な計算機設備を増強した。

D. 考察

1. 全ゲノム相関解析を実施する情報分析プログラムの研究開発を終え、今後は疾患関連変異を探索する。

2. 1分子リアルタイムシーケンサーの解読精度が平成26年度に著しく改善し、平均長が5000塩基から1万塩基を超えるようになった。そのため検出可能な大規模構造変異(短鎖繰返配列を含む)の範囲は広がるであろう。
3. 計算機設備の増強は今後も継続する必要がある。特に二次記憶装置はデータ解析量に比例して追加の必要がある。

E. 結論

当初の計画通り、計算機設備を整備し、全ゲノム相関解析をエキソームへと拡大し、低頻度の疾患関連変異を検出するための理論的考察を終え検出ソフトウェアを研究開発した。さらに短鎖繰返配列および大規模構造変異を検出するためのソフトウェアを研究開発した。

F. 健康危険情報

該当せず。

G. 研究発表

1. 論文発表

- Koichiro Doi, Taku Monjo, Pham H Hoang, Jun Yoshimura, Hideaki Yurino, Jun Mitsui, Hiroyuki Ishiura, Yuji Takahashi, Yaeko Ichikawa, Jun Goto, Shoji Tsuji, Shinichi Morishita. Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. *Bioinformatics*. 30(6), 815-822 (2014)
- Kazuki Ichikawa, Shinichi Morishita. A simple but powerful heuristic method for accelerating k-means clustering of large-scale data in life science.

- IEEE/ACM Transactions on Computational Biology and Bioinformatics. 11(4), 681-692 (2014)
- Kishin Koh, Hiroyuki Ishiura, Michiaki Miwa, Jun Yoshimura, Jun Mitsui, Jun Goto, Shinichi Morishita, Shoji Tsuji, Yoshihisa Takiyama. Exome sequencing shows a novel de novo mutation in ATL1. *Neurology and Clinical Neuroscience*. 2(1), 1-4 (2014)

2. 学会発表

該当せず。

H. 知的財産権の出願・登録状況(予定を含む)

1. 特許取得

該当せず。

2. 実用新案登録

該当せず。

3. その他

該当せず。

厚生労働科学研究委託費(難治性疾患実用化研究事業)

委託業務成果報告(業務項目)

神経疾患の集中的な遺伝子解析及び原因究明に関する研究
(遺伝性神経疾患の病因遺伝子の解明)

担当責任者 石浦 浩之
東京大学医学部附属病院神経内科助教

研究要旨

発症機構が未解明の、遺伝性神経疾患に対して、高精度のゲノム配列解析と情報解析を行い、原因遺伝子同定を目指した。具体的には、家族性筋萎縮性側索硬化症、遺伝性痙性対麻痺、家族性多系統萎縮症を中心にゲノム DNA 収集を行い、またエクソーム解析・全ゲノム配列解析を行った。今後、情報解析により見出された変異の機能解析や、候補遺伝子についての validation を進める。

A. 研究目的

本研究は、神経疾患(難病)を対象疾患群として、1. 次世代シーケンサーを駆使したゲノム配列解析拠点の整備、2. 次世代シーケンサーから得られる膨大な規模のゲノム配列情報を処理するゲノムインフォマティクス拠点の整備により、発症機構が未解明の数多くの神経疾患について、発症に関与する遺伝子を解明し、分子病態機序に基づく治療法開発研究を実現することを目的としている。

中でも遺伝性神経疾患の解明に向けて、小家系、家系数が少ないなどの制限があることが多いが、網羅的なゲノム配列解析、変異解釈の高機能化により、病因遺伝子の発見を加速させることができるのではないかと考え、研究を行った。

B. 研究方法

主に家族性筋萎縮性側索硬化症、遺伝性痙性対麻痺、家族性多系統萎縮症家系に関して、家系収集を進め、ハイスループット連鎖解析システムにあわせ、エクソーム解析・全ゲノム配列解析を施行し、原因変異同定を目指した。

エクソーム解析は Agilent 社の SureSelect V5+UTR キットを用い、HiSeq2500 を用いて 101 塩基ペアエンド法で塩基配列解析を行った。全ゲノム配列解析は、PCR-free TruSeq DNA sample prep kit を用いてサンプル調整を行い、HiSeq2500 を用いて塩基配列解析を行った。短鎖長配列は、BWA を用いて参照配列 hg19 に alignment を行い、SAMtools を用いて variant call を行った。現在まで蓄積されてきた対照者のエクソーム解析・全ゲノム配列

解析

データにあわせ、今年度新規に解析されたコントロール例を合わせて、対照者とした。また、変異の minor allele frequency については、各種公開データベースを用いた。

(倫理面への配慮)

ヒトゲノム・遺伝子解析研究に関する倫理指針に基づいて研究を行った。

C. 研究結果

1. 家族性筋萎縮性側索硬化症

本年度、新たに5家系の発症者の家族性筋萎縮性側索硬化症症例のゲノムDNAを収集し、エクソーム解析を行った。

家族性筋萎縮性側索硬化症としては、総計68家系となり、うち *SOD1* 27例、*FUS* 7例、*TARDBP* 2例と、本邦においてはこの3つの遺伝子の変異が比較的多く認められることが判明した。その他、*ERBB4*、*C9ORF72* に変異を1例ずつ認めた。欧米で多い *C9ORF72* 内の6塩基リピート伸長変異は1例と、本邦における頻度は低いと考えられた。

その他、稀ながら、他の遺伝子の新規遺伝子変異の候補が2家系(*VCP*、*HNRNPA1*)で認められた。これらの変異については、未だ分子病態が明らかとなっていない点があり、現在機能解析を行っている。

2. 遺伝性痙性対麻痺

家族歴より、常染色体劣性遺伝が疑われた症例と、孤発例でも複合型である症例で変異が認められていない144症例に関して、エクソーム解析を行った結果、58例(約40%)については既知の遺伝子に変異が認められた。変異が同定されなかった症例に関して、rare

variantを2アレル以上有する遺伝子をピックアップした。このような2アレル以上変異を有する遺伝子のうち、複数症例で共通して認められるのが13個あり、12個は2家系で共通するのみであったが、1個については4家系で共通する可能性が示唆された。今後、この遺伝子について、validationを進める。共分離の確認や他の家系においてさらに変異が同定されるかどうか検討する必要があるものの、遺伝性痙性対麻痺の新規原因遺伝子の候補となると考えられた。

3. 家族性多系統萎縮症

今まで、稀ではあるが家族性多系統萎縮症の解析から *COQ2* 遺伝子変異を見出し、本遺伝子は孤発性においてもリスクとなることが判明している(Mitsui et al. 2013 *NEJM*)。このことより、引き続き家族例の収集を進めることが、孤発性疾患のリスク遺伝子発見にもなりうると考え、収集を続けている。

本年度は多発家系1家系について新たな発症者を認め、新規発症者2名についてゲノムDNAを採取の上、全ゲノム配列解析を行った。

また、新規に多系統萎縮症家系2家系(発症者3、非発症者3名)を収集し、全ゲノム配列解析を行った。

また、孤発であるものの、両親に近親婚があり、常染色体劣性遺伝の可能性のある20家系について、エクソーム解析に合わせ、Affymetrix Genome-Wide SNP array 6.0で遺伝子型を決定し、ハイスループット連鎖解析を行い、原因遺伝子同定を目指している。

D. 考察

家族性筋萎縮性側索硬化症、遺伝性痙性対麻痺、家族性多系統萎縮症を中心にエクソーム解析、全ゲノム配列解析を行い、必要に応じてハイスループット連鎖解析を併用し、遺伝性神経疾患の原因遺伝子同定に向けて検討を行った。

遺伝性痙性対麻痺の検討をとってみても、既知の原因遺伝子に変異が見出される例が未だ半分以下であり、原因遺伝子を数多く同定することは、疾患のさらなる理解に有用で、また治療法開発の有益なシーズになる可能性がある。

引き続き、家系収集を続け、原因遺伝子の同定をめざす。

E. 結論

高精度な情報解析、豊富な日本人対照者のデータを用いて、遺伝性神経疾患について解析を行った。いくつか貴重なデータが出つつあり、機能解析・家系収集を通して引き続き検討を行う。

F. 健康危険情報

該当せず。

G. 研究発表

1. 論文発表

Koh K, Kobayashi F, Miwa M, Shindo K, Isozaki E, Ishiura H, Tsuji S, Takiyama Y.
Novel mutations in the PNPLA6 gene in Boucher-Neuhauser syndrome. *J Hum Genet* 2015 Epub ahead of print.

2. 学会発表

H. Ishiura, K. Koh, H. Shimazaki, J. Mitsui, Y. Takahashi, J. Goto, K. Yoshimura, K. Doi, S.

Morishita, H. Sasaki, Y. Takiyama, S. Tsuji, JASPAC (Japan Spastic Paraplegia Research Consortium). Exome analysis of 116 patients supposed to be autosomal recessive hereditary spastic paraplegia established molecular diagnoses of 49 patients with broad genetic heterogeneities. The 64th annual meeting of American Society of Human Genetics 2014. San Diego, USA.

H. 知的財産権の出願・登録状況(予定を含む)

1. 特許取得

該当せず。

2. 実用新案登録

該当せず。

3. その他

該当せず。

厚生労働科学研究委託費(難治性疾患実用化研究事業)

委託業務成果報告(業務項目)

孤発性神経疾患の病因遺伝子の解明

担当責任者 三井 純

東京大学医学部附属病院神経内科助教

研究要旨

代表的な孤発性の神経変性疾患である多系統萎縮症, 筋萎縮性側索硬化症の背景にある遺伝因子を解明するため, エクソーム解析を行った. 多系統萎縮症においては, 新規の疾患感受性遺伝子として GBA 遺伝子を初めて同定した. 筋萎縮性側索硬化症においては, 一見孤発性と考えられる患者の約 3.1% に遺伝性筋萎縮性側索硬化症の病因遺伝子変異を認めた.

A. 研究目的

多系統萎縮症, 筋萎縮性側索硬化症は, その大部分の患者は孤発性に発症し, その病因がほとんど分かっていない. これらの疾患の背景にある遺伝因子を解明するため, エクソーム解析を行った.

B. 研究方法

1) 国際多施設共同研究で収集された孤発性多系統萎縮症患者 969 例(日本 574 例, 欧州 223 例, 北米 172 例), 健常対照者 1509 例(日本 900 例, 欧州 315 例, 北米 294 例)に対して, エクソーム解析を含む網羅的なゲノム解析を行った. 既に解析を終えている多発家系例の rare variants の検討などから, 関連遺伝子の候補として GBA 遺伝子が挙がっていた. GBA 遺伝子の変異を網羅的に調べ機能障害性と考えられる変異(ゴーシェ病患

者で特異的に認められる変異)について集計して検定を行った.

2) 日本人孤発性筋萎縮性側索硬化症 319 例に対して, エクソーム解析を行った. 孤発性筋萎縮性側索硬化症の発症者の中に, 遺伝性筋萎縮性側索硬化症(常染色体優性遺伝)の病原性変異を持つ例がな

いか検索した.

(倫理面への配慮)
ヒトゲノム・遺伝子解析研究に関する倫理指針に基づいて研究を行った.

C. 研究結果

1) 日本, 欧州, 北米の 3 シリーズにおける患者・対照群の GBA 遺伝子機能障害性変異のキャリア頻度についてメタ解析を行ったところ, オッズ比 2.44 (95%信頼区間 1.14-5.21), p 値 0.029 にて, GBA 遺伝子機能障害性変異は多系統萎縮症患

者と関連した。サブタイプ別では、小脳症状優位型(MSA-C)において強い関連(p値 0.0073)が得られた。パーキンソニズム優位型(MSA-P)においては、GBA 遺伝子機能障害性変異のキャリア頻度が高い傾向を認めたが、有意差は得られなかった。

- 2) 日本人孤発性筋萎縮性側索硬化症 319 例のうち、9 例(3.1%)に遺伝性筋萎縮性側索硬化症(常染色体優性遺伝)の病原性変異を認めた(SOD1 遺伝子変異 4 例, TARDBP 遺伝子変異 2 例, FUS 遺伝子変異 2 例, C9ORF72 遺伝子変異 1 例)。

D. 考察

孤発性の神経変性疾患である多系統萎縮症、筋萎縮性側索硬化症に対して、大規模で網羅的なゲノム解析を行った。エクソームのデータから、ゲノムワイドに信頼性の高い疾患感受性変異・遺伝子の検索を行うためには、サンプル規模のさらなる拡大が必要である。一方、サンプル規模が小さい段階でも、多発家系例に注目したアプローチ、候補遺伝子アプローチなどで、効率よく疾患感受性遺伝子を同定できる可能性がある。

- 1) 多系統萎縮症の多発家系例の解析によって得られた、複数の発症者で共有される rare variants のうち、多系統萎縮症の病態との関連が想定される遺伝子を抽出した。今回は、GBA 遺伝子に注目し、現時点の規模における p 値はマージナルであるが、有意差を得た。今後も検討を続ける。
- 2) 日本人孤発性筋萎縮性側索硬化症 319
- 3) 例を検索し、うち 3.1%は遺伝性筋萎縮性側索硬化症(常染色体優性遺伝)であっ

た。今後、遺伝性筋萎縮性側索硬化症(常染色体劣性遺伝)の変異キャリア頻度の検討を含め、候補遺伝子アプローチにて疾患感受性遺伝子を検索していく。

E. 結論

多系統萎縮症、筋萎縮性側索硬化症の背景にある遺伝因子を解明するため、エクソーム解析を行った。多系統萎縮症においては、新規の疾患感受性遺伝子として GBA 遺伝子を初めて同定した。筋萎縮性側索硬化症においては、一見孤発性と考えられる患者の約 3.1%に遺伝性筋萎縮性側索硬化症の病因遺伝子変異を認めた。

F. 健康危険情報

該当せず。

G. 研究発表

1. 論文発表

Mitsui J, et al. Variants associated with Gaucher disease in multiple system atrophy. *Ann Clin Transl Neurol.* 2015 (DOI: 10.1002/acn3.185)

2. 学会発表

該当せず。

H. 知的財産権の出願・登録状況(予定を含む)

1. 特許取得 該当せず。
2. 実用新案登録 該当せず。
3. その他 該当せず。

厚生労働科学研究委託費(難治性疾患実用化研究事業)

委託業務成果報告(業務項目)

神経疾患の病態機序の解明

担当責任者 辻 省次
東京大学医学部附属病院神経内科教授

研究要旨

近年、神経筋変性疾患の病態に RNA 異常が関わっていることを示唆する報告が増えている。特に筋萎縮性側索硬化症については原因として RNA 異常を示唆する報告が増えており、RNA processing の解析より病態機序を解明することを目的とし、次世代シーケンサーを用いた網羅的な RNA 配列解析(RNA-seq)のパイプラインの構築を行った。今後、多数の症例の検討を行い、病態機序の解明を目指す。

A. 研究目的

近年、特に筋萎縮性側索硬化症について、RNA の関与が指摘されるようになってきている。家族性筋萎縮性側索硬化症の原因遺伝子として同定された *TARDBP* と *FUS* にはともに RNA 結合ドメインが存在しており、病態に RNA 代謝異常が存在するのではないかと示唆されるようになった。また 2011 年に家族性筋萎縮性側索硬化症・前頭側頭型認知症の原因として *C9orf72* 内のイントロンにある GGGGCC という 6 塩基反復配列の伸長が認められたが、これは非翻訳領域に存在し、RNA に転写された上で RNA foci と呼ばれる封入体を脊髄運動ニューロンなどで作ることが指摘されている。他にも、脊髄小脳変性症、脆弱 X 関連振戦失調症候群、筋強直性ジストロフィーなどにおいてもイントロン・非翻訳領域のリピート伸長変異が

見出されており、これら由来の異常 RNA が神経筋変性疾患の病態に強く関わっていることが推測される。今回、RNA processing の観点より神経疾患の病態機序を解明するために、RNA 配列解析を行った。

B. 研究方法

網羅的な RNA 配列解析を行うために、HiSeq2500 を用いて、RNA-seq を行った。具体的には、Trizol を用いて total RNA を抽出した後、Bioanalyzer を用いて RNA の品質チェックを行った。RNA については、Ribo Zero gold を用いて rRNA とミトコンドリア tRNA の除去を行った後、TruSeq RNA sample prep kit v2 もしくは TruSeq Stranded Total RNA with Ribo-Zero Gold kit を用いてサンプル調整を行った。逆転写反応を行う際に、poly T と random primer の 2 種類について検討を行った。

HiSeq2500 を用いて、サンプルを 101 塩基ペアエンド法で塩基配列解析を行い、得られた短鎖長配列については、TopHat を用いて、参照配列 hg19 と、参照トランスクリプトーム (gencode v14) を用いて alignment を行った。各遺伝子の発現量、アイソフォーム解析については Cufflinks を用いて行い、データの可視化には CummeRbund を用いた。

(倫理面への配慮)

ヒトゲノム・遺伝子解析に関する倫理指針に基づいて得られたサンプルについて、研究を行った。

C. 研究結果

1. 対照者の脳の RNA-seq

まず、RNA-seq に用いる方法を検討するために、逆転写反応を poly T で行った場合と、random primer で行った場合を比較した。対照者 2 名の後頭葉から抽出された RNA について、両者で逆転写を行い、TruSeq RNA sample prep kit v2 を用いて解析を行った。結果、poly T を用いた反応では cDNA の 3' 側のカバレッジは良好であるものの、長い cDNA については 5' 側のカバレッジが低値となってしまうことが判明した。Random primer を用いて逆転写反応を行った場合は、比較的カバレッジが均一となり、解析に有用と考えられた。

2. 対照者の他組織の RNA-seq

上記のデータは、脳における各遺伝子の発現量の基礎データとなると考えられた。今後の解析を念頭に、その他の組織・細胞の RNA-seq データを取得しておくことは有用と考えられ、対照者の後頭葉 6 例 (合計 8 名)、対

照者リンパ芽球様細胞 2 例、対照者肝臓 3 名について、random primer と Ribo Zero Gold、TruSeq RNA sample prep kit v2 を用いて RNA-seq を行った。Alignment を行い、各組織・細胞における遺伝子発現量を検討する上でのデータとした。

3. 筋萎縮性側索硬化症患者大脳・小脳の RNA-seq

筋萎縮性側索硬化症患者における RNA processing 異常を検出するために、脊髄組織を用いた RNA-seq を検討していたが、組織量が極めて希少であるため、まずは大脳における RNA 代謝異常を検出することを目的に、大脳 (n=3)、小脳 (n=2) を用いて RNA-seq を行った。対照として、1. で得られたデータを用いて解析を行った。結果大脳のサンプルにおいて、27 個の有意に発現変動が認められる遺伝子を見出した。これらの変動遺伝子を確認するために、さらに症例数を増やして RNA-seq を行うとともに、今後脊髄の RNA-seq を行うことを検討していく。

また、RNA 配列中の反復配列を検出するプログラムを作成しており、その有用性についても今後検討を行う。

D. 考察

剖検組織を用いて、RNA-seq を行い、網羅的に RNA 配列を検討した。パイプラインとして、random primer を用いた逆転写反応を選択し、情報解析としては TopHat、Cufflinks、CummeRbund を用いた。RNA の方向性を維持した解析 (Stranded) を行うかどうかは状況によるが、対照 RNA-seq データが増えていけば、今後は Stranded RNA-seq が主流になると考えられた。