

## 長寿医療研究開発費 平成 29 年度 総括研究報告

大規模ゲノム-RNA 解析による網羅的な老年病原因、感受性分子の同定と解析 (29-45)

主任研究者 尾崎浩一 国立長寿医療研究センター 臨床ゲノム解析推進部 (部長)

### 研究要旨

老年病の中でも認知症の患者数は全世界で増加の一途をたどっており、本邦においてもその患者数は 500 万人 (2013 年、厚生労働省研究班推計) に達する勢いである。大部分の認知症は、糖尿病や虚血性心疾患と同様に生活習慣病と捉えることができ、環境因子と遺伝因子が複雑に絡み合っ て発症すると考えられるが、これまでの双子疫学研究による認知症、特にアルツハイマー病 (AD) の発症に与える遺伝因子の割合は 58%~79% であることが証明されており、その大部分を遺伝因子が占めていることが明らかとなっている。したがって、この遺伝因子群を同定し、その役割を精査することから疾患の分子メカニズムが解明でき、エビデンスに基づく予防法や治療法の開発に大きく貢献できると考えられる。

### 主任研究者

尾崎浩一 国立長寿医療研究センター 臨床ゲノム解析推進部 (部長)

### 分担研究者

新飯田俊平 国立長寿医療研究センター メディカルゲノムセンター (センター長)

伊藤 薫 理化学研究所 循環器疾患研究チーム (チームリーダー)

### A. 研究目的

長寿医療研究センター (NCGG) バイオバンク等の試料 (血液や組織より得られた DNA、RNA) を活用した日本人に特化したジェノタイピングプラットフォームによる 10,000 人規模のゲノムワイド関連解析 (GWAS) データと次世代シーケンサーによる全 RNA 配列 (最終的には 1000 検体を目標とする) データ (スプライスバリエントも含めた全 RNA の正確な発現情報、バリエーション情報、ハプロタイプフェージング等) を遺伝統計学的に統合、解析することによって認知症等の老年病の原因、感受性分子を網羅的に探索する。これらの統合データから疾患パスマーカーの同定、解析や遺伝的リスクスコア等の算出、ドラッグリポジショニング、さら

に機械学習や AI も取り入れることにより疾患判別モデルの開発等からエビデンスに基づく新規診断法、治療薬の開発に応用する。

## B. 研究方法

日本人に特化した全ゲノムジェノタイピングアレイであるジャポニカアレイ（株式会社東芝）による一塩基多型（SNP）ジェノタイピングを進めた。1000ゲノムデータ（[www.internationalgenome.org/](http://www.internationalgenome.org/)）をリファレンスとしたインピュテーション法による疑似全ゲノム配列を取得し、統計学的な解析によりクオリティフィルターを通してデータをクリーンアップし、年齢、性別を共変数としたロジステック回帰分析によるゲノムワイド関連解析（GWAS）を施行した。一方で同一検体について、バフィーコートよりRNAを抽出し、RNAのクオリティを確保できた検体について全RNAライブラリを構築した。全RNA配列決定は（株）ジーンウィズ社に外注した。

また、RNA シークエンスのドライ解析パイプラインの確立については、多くは遺伝子の発現解析にとどまっているが、スプライス異常、エキソン利用率の変化、イントロン保持、アレル特異的発現、転写産物の変化など様々な現象が観察できる。しかしながら現在の解析技術ではそれぞれ断片的に解析可能であるが、1) 確立したパイプラインが存在しない、2) DNA 変異との関連性を検討したものが少ないという欠点が挙げられる。従って、この2つについて技術的な課題を解決しつつ、RNA シークエンスデータの解析を行う必要がある。本年度は技術的な問題を解決するため 1) スプライシングバリエーションの DNA 変異からの予測と RNA シークエンス結果からの検出方法の確立、2) イントロン保持の検出についてアルゴリズムの検証と実際応用にあたってパイプラインの設計をパブリックデータとマウスのデータを用いて行った。

### （倫理面への配慮）

本研究は、「人を対象とする医学系研究に関する倫理指針」および「ヒトゲノム・遺伝子解析研究に関する倫理指針」に則り、国立研究開発法人国立長寿医療研究センター倫理・利益相反委員会の承認を得て行われた。本年度における理化学研究所の分担研究については、RNA 解析方法論の確立についてパブリックデータ、マウスデータを用いた *in silico* 解析を行っているため、倫理面の問題は生じない。

## C. 研究結果

国立長寿医療研究センターバイオバンクが保有する DNA サンプルを用い、日本人に特化したジャポニカアレイによる一塩基多型（SNP）の全ジェノタイピングデータおよびバフィーコートによる全 RNA 配列データの取得を行ってきた。現在までに約 7,100 例（ア

ルツハイマー病；AD 2,391 例、認知機能正常群；Control 3,346 例、軽度認知障害；MCI 601 例、血管性認知症；VaD 88 例、前頭側頭型認知症；FTD 25 例、レビー小体認知症；DLB 128 例、その他 553 例）の全ゲノムジェノタイピングデータを取得している。ここで得られたジェノタイプデータから、試験的に AD 2,250 例、コントロール 1,848 例を用いて、1000 ゲノムデータ ([www.internationalgenome.org/](http://www.internationalgenome.org/)) を用いたインピュテーションを施行し、GWAS を施行したところ、既報通り *APOE* 座位において非常に有意な AD との関連が認められると同時に、集団階層化の指標である  $\lambda_{GC}$  が 1.00 となり、集団の階層化がほとんど存在しないことが示唆された。また、本試験的 GWAS においては新規の GWAS 有意性 ( $P = 5 \times 10^{-8}$  未満) を示す染色体領域は見られなかったが、示唆的な有意性 ( $P = 10^{-5}$  未満) を示す座位が第 3、5、8、10、13、15、18 番染色体などに認められ、これらの中には低頻度かつ東アジア人にしかアレル頻度が認められない SNP も存在していた。また、これまでに全世界から GWAS 等で報告された既存の AD 関連座位の中で、今回の日本人集団でも再検証できた座位としては、8 番染色体の *CLU* 遺伝子、11 番染色体の *PICALM1* と *SORL1*、19 番染色体の *ABCA7* があげられる ( $P < 0.005$ )。

全 RNA 配列の取得については、約 200 例について高品質な RNA を取得し、全 RNA 配列解析用のライブラリ作製を進めてきた。全 RNA の質の指標である RNA integrity number (*RIN*) は概ね 7.0 以上が得られており、RNA ライブラリの質も良好であった。次世代シーケンサーによる配列決定およびそのクオリティーチェックにおいて問題ないことを確認している。

また、RNA シークエンスのパイプライン開発については、スプライシングバリエーションとなる DNA 変異を検出するアルゴリズムは幾つか存在するが、それぞれ費やすコンピュータ時間と精度について検証したところ、

- 1) 古典的アルゴリズムである MaxEntScan は計算が非常に素早かった
- 2) 深層学習アルゴリズムを用いた HAL (Cell 2015) や SPINDEX (Science 2014) などは精度も従来のものより良好であるが非常に計算が遅かった。

以上より多数の DNA 変異を解析することを考慮し 1) を採用した。

また RNA シークエンスによりスプライシングバリエーションの検証であるが、よく用いられている RNA シークエンスアライナーである TOPHAT2 はスプライシング検出の正確性に問題があることが発覚したため、STAR を用いてアライメントを行うこととした。この過程において DNA 変異 → RNA スプライシング異常を紐づけて予測、検証するパイプラインを確立した。イントロン保持に関してはいくつかのアルゴリズムの性能を検証した。

#### D. 考察と結論

日本人に特化したジャポニカアレイによる 7,132 例のジェノタイピングを施行し、試験的な GWAS をアルツハイマー病に対して行ったところ、非常に良好な  $\lambda$  値を示し、本施設により収集されたサンプルが階層化なく均一なサンプルでありことが確認できた。

GWAS 有意性を得ることのできた新規座位は存在していないが、サンプル数の増加、インピュテーション解析を東北メディカルメガバンクの保有する一般人 3,500 人の全ゲノム配列データにより構築されたリファレンスパネルを使用することにより新規かつ低頻度なバリエーションが GWAS 有意性として挙がってくるものと考えられる。

全 RNA 配列解析については、200 例について次世代シーケンサーによる配列決定を完了し、その質も問題ないことを確認しているため、順次サンプル数を増加し、統合解析に耐えうるサンプル数の蓄積を目指す。

ドライでの RNA 解析パイプライン構築については、スプライシングバリエーションの検出として MaxEntScan による予測と ENSEMBL データベースによる効果の予測、STAR によるスプライシングギャップの検証が最も能率がよく、これらを用いたパイプラインを確立している。イントロン保持については開発途中であり、更なる検証と確認が求められる。

#### E. 健康危険情報

なし。

#### F. 研究発表

##### 1. 論文発表

平成 29 年度

- 1) Nagata Y, Hirayama A, Ikeda S, Shirahata A, Shoji F, Maruyama M, Kayano M, Bundo M, Hattori K, Yoshida S, Goto Y, Urakami K, Soga T, Ozaki K and Niida S. Comparative analysis of cerebrospinal fluid metabolites in Alzheimer's disease and idiopathic normal pressure hydrocephalus in a Japanese cohort. *Biomarker Res.* Jan 22;6:5, 2018.
- 2) Sudo T, Okada Y, Ozaki K, Urayama K, Kanai M, Kobayashi H, Gokyu M, Izumi Y, Tanaka T. Mutations in *NOD2* cause aggressive periodontitis. *Journal of Dental Research* 96(10):1100-1105 (2017).
- 3) Ozaki K. Molecular genetics of coronary artery disease. *eLS (Encyclopedia of Life Sciences)* November 2017. DOI: 10.1002/9780470015902.a0027329
- 4) Ito K, Patel PN, Gorham JM, McDonough B, DePalma SR, Adler EE, Lam L, MacRae CA, Mohiuddin SM, Fatkin D, Seidman CE, Seidman JG. Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. *Proc Natl Acad Sci U S A.* 114:7689-7694, 2017.
- 5) Low SK, Takahashi A, Ebana Y, Ozaki K, Christophersen IE, Ellinor PT; AFGen Consortium, Ogishima S, Yamamoto M, Satoh M, Sasaki M, Yamaji T, Iwasaki M, Tsugane S, Tanaka K, Naito M, Wakai K, Tanaka H, Furukawa T, Kubo M, Ito K\*,

Kamatani Y\*, Tanaka T\*. Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat Genet.* 49:953-958, 2017.

\*Co-corresponding author.

※発表誌名、巻号・頁・発行年等も記載すること。

## 2. 学会発表

平成 29 年度

- 1) Ozaki K. Biological and pharmaceutical investigation of molecules associated with the increased risk coronary artery disease. *Genomic Medicine* 2018, Houston (February 28, 2018)
- 2) 尾崎浩一、アルツハイマー病のゲノム解析 日本医療研究開発機構主催 脳と心の研究会 講演（東京、2017年11月）
- 3) 尾崎浩一、孤発性アルツハイマー病のゲノム解析 第59回 日本老年医学会学術集会 シンポジウム 講演（名古屋、2017年6月）
- 4) 松永紘、伊藤薫 「全ゲノムシーケンスを用いた心筋梗塞メカニズムの解明」 *Molecular Cardiovascular Metabolic Conference*（大阪、2017年9月）
- 5) 伊藤薫 「Cutting Edge of Genomic Research:GWAS, NGS, and Others」第64回 日本不整脈心電図学会学術大会（横浜、2017年9月）

## G. 知的財産権の出願・登録状況

### 1. 特許取得

なし

### 2. 実用新案登録

なし

### 3. その他

なし

※予定を含めて記載すること。該当がない場合には「なし」と記載すること。