

「ゲノムデータ等を用いた疾患関連因子の網羅的探索と疾患発症リスク予測モデルの開発」  
に関する研究（30-29）

主任研究者 重水 大智 国立長寿医療研究センター MGC 臨床ゲノム解析推進部  
(ユニット長)

#### 研究要旨

本研究は、臨床情報と次世代シーケンサー（NGS）から得られる大量のゲノム配列データとオミックスデータを用い、①認知症等の疾患発症リスク遺伝子の同定、および②疾患発症リスク予測モデルの開発を実施する。研究対象疾患として、アルツハイマー型認知症（AD）、脳血管性認知症（VaD）、レビー小体型認知症（DLB）のほか、ゲノムデータが蓄積され、未解析の疾患も対象とする。本研究の最終目的は、疾患発症リスク予測モデルの開発により疾患の早期発見、早期予防に貢献することである。

#### 主任研究者

重水 大智 国立長寿医療研究センター MGC 臨床ゲノム解析推進部 (ユニット長)

#### 分担研究者

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

秋山 真太郎 国立長寿医療研究センター MGC 臨床ゲノム解析推進部 (研究員)

## A. 研究目的

認知症の高齢者は毎年増加傾向にあり、2025年にはその数は470万人に達すると言われている。しかしながら、その病態解明は未だ不明であり、根本的治療薬や予防法も十分に確立されていない。したがって、認知症の病態解明を含めた認知症の早期発見や診断法の確立が急務である。

近年、ヒトゲノム全体の一塩基多型(SNP)をジェノタイピングし、疾患群と健常者群を比較することで、疾患群に特異的な疾患発症リスク因子を網羅的に探索するゲノムワイド関連解析(Genome-Wide Association Study: GWAS)が盛んに行われ、様々な疾患でリスク因子が同定されている。またそのリスク因子を用いた疾患の発症予測の研究も行われている。二型糖尿病やリウマチに関する報告は多数あるが、アルツハイマー型認知症における報告はほとんどない。そこで本研究の目的は、アルツハイマー型認知症発症者と非発症者のゲノム情報等から疾患発症リスク因子または遺伝子を網羅的に探索し、そのリスク因子をもとに疾患発症予測モデルの開発を行う。日本人を基盤とした認知症の早期発見、早期予防への貢献が期待される研究は独創的かつ画期的である。このアプローチはさまざまな疾患への応用が可能である。

## B. 研究方法

具体的には次のようにアプローチする。1)疾患群(ケース群)と健常者群(コントロール群)の大規模なゲノム配列データ等を比較することで(Case-Control study)、疾患群において統計学的に有意に多い変異(疾患関連因子)を網羅的に探索、同定する。2)その因子と発現データとの関連性を調べることで疾患関連因子に制御される遺伝子群(疾患発症リスク遺伝子)を網羅的に同定する。3)同定された遺伝子群に対して機械学習(またはAI)の手法を適応させることで最適な疾患発症リスク予測モデルの開発を実施する。4)開発されたモデルの検証には、後ろ向きデータによる交差検証に加え、前向きコホートデータにて評価する。例えばMCIの患者が今後どの認知症に移行する可能性が高いかなどをこの予測モデルを用いて長期的に調べる。

### (倫理面への配慮)

本研究で使用する全ての検体は、解析に対する同意を書面で得ている。また、該当検体を使用した研究については国立長寿医療研究センター、理化学研究所、東京医科歯科大学における倫理委員会の承認を得て実施する。

### C. 研究結果

現在、AD、VaD、DLB のオミクスデータ(全ゲノムシーケンスデータ、全エクソームシーケンスデータ、RNA シーケンスデータ、マイクロ RNA 発現データ)を測定している。データの測定は進行中であるが、先行してマイクロ RNA(miRNA)の発現データから認知症の発症リスクマーカーを同定し、疾患発症リスク予測モデルの構築を行った。

今回の解析に使用した検体数は 1,569 検体である。内訳は AD が 1,021 検体、VaD が 91 検体、DLB が 169 検体、健常者が 288 検体である。病態ごとにデータを学習群と検証群に分け、病態の学習群と健常者群を比較することで疾患に關与するバイオマーカーの選定、予測モデルの構築を機械学習手法のひとつである **Supervised PCA logistic method** を用いて実施した。構築した予測モデルの精度評価は独立な検証群で行い、AD で 78 個、VaD で 86 個、DLB で 110 個の miRNA が各病態を選別する有用なバイオマーカーとして選定され、その予測能は AD で AUC が 0.874、VaD で 0.867、DLB で 0.870 と高精度であった(上記バイオマーカーは特許出願: 特願 2018-085652)。この予測モデルはまた前向きデータにも適応され、MCI から AD へのコンバートの予測も高精度に行えることが実証された。これらの解析結果は英国科学雑誌 *Communications Biology* に掲載され、AMED からプレスリリースされ、さまざまなメディアにも取り上げられた([https://www.amed.go.jp/news/release\\_20190225-03.html](https://www.amed.go.jp/news/release_20190225-03.html))。現在、その他の機械学習の手法を適応させることで、さらに予測能の改善が DLB に対して認められたため、それらの結果を英国科学雑誌 *Scientific Reports* に投稿中である。

### D. 考察と結論

現在、miRNA 単独のデータからバイオマーカーの選定、疾患予測モデルの構築を行った。今後検体数が増え、またゲノム配列データ、遺伝子発現データ等のオミクスデータを加味することで、より高精度な疾患発症予測モデルに必要なバイオマーカーの選定、モデル構築が可能になると考えられる。

### E. 健康危険情報

なし

### F. 研究発表

#### 1. 論文発表 (\*: 責任著者)

1. **Shigemizu, D.\***, Akiyama, S., Asanomi, Y., Boroevich, KA, Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., Ochiya, T., and

- Niida, S. (2019) Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. *Commun Biol* Feb.25; 2:77.
2. Reddy, HM., Sharma, A., Dehzangi, A., **Shigemizu, D.**, Chandra, A., and Tsunoda, T. (2018) GlyStruct: Glycation prediction using structural properties of amino acid residues. *BMC Bioinformatics* 19(Suppl 13):547.
  3. Yamaguchi-Kabata, Y., Morihara, T., Ohara, T., Ninomiya, T., Takahashi, A., Akatsu, H., Hashizume, Y., Hayashi, N., **Shigemizu, D.**, Boroevich, KA., Ikeda, M., Kubo, M., Takeda, M., and Tsunoda, T. (2018) Integrated analysis of human genetic association study and mouse transcriptome suggests LBH and SHF genes as novel susceptible genes for amyloid- $\beta$  accumulation in Alzheimer's disease. *Human Genetics* Jul. 13.
  4. López, Y., Kamola, PJ., Sharma, R., **Shigemizu, D.**, Tsunoda, T., and Sharma, A (2018) Computational pipelines and workflows in Bioinformatics. *Encyclopedia of Bioinformatics and Computational Biology* Sep 1.
  5. Nishino, J., Kochi, Y., **Shigemizu, D.**, Kato, M., Ikari, K., Ochi, H., Noma, H., Matsui, K., Morizono, T., Boroevich, K., Tsunoda, T., and Matsui, S. (2018) Empirical Bayes estimation of semi-parametric hierarchical mixture models for unbiased characterization of polygenic disease architectures. *Frontiers in Genetics* Apr 24.
  6. **Shigemizu, D.\***, Miya, F., Akiyama, S., Okuda, S., Boroevich, K., Fujimoto, A., Nakagawa, H., Ozaki, K., Niida, S., Kanemura, Y., Okamoto, N., Saitoh, S., Kato, M., Yamasaki, M., Matsunaga, T., Mutai, H., Kosaki, K., and Tsunoda, T. (2018) IMSindel: An accurate intermediate-size indel detection tool incorporating de novo assembly and gapped global-local alignment with split read analysis. *Sci Rep* Apr 4;8(1):5608.

## 2. 学会発表

(招待講演)

1. **重水 大智:** 「ヒトゲノム・オミクスデータによる疾患原因変異の同定と疾患発症予測モデル開発」名古屋市立大学医学部セミナー, 名古屋, Nov. 27, 2018.
2. **重水 大智:** 「ヒトゲノム・オミクスデータによる疾患原因変異の同定と疾患発症予測モデル開発」九州工業大学セミナー, 福岡, Nov. 15, 2018.
3. **Shigemizu, D.:** 「ゲノムデータを用いた疾患原因変異の同定と疾患発症予測モデル開発の現在」第7回生命医薬情報学連合大会, 鶴岡, Sep 20, 2018.
4. **Shigemizu, D.:** 「AI-Aided Genomic Studies for Disease Prediction」2018 AI Future: Application on Geriatrics Care Symposium, Taipei, July 23, 2018.

(口頭発表)

1. 芦崎 晃一, 山本 幸二, 川崎 洋, 川上 英良, **重水 大智**, 桜田 一洋, 角田 達彦, 海老原 全, 天谷 雅行: 医療データ駆動型研究を推進する医療データ収集システムの構築と導入効果、**第38回日本医療情報学連合大会, 2018**, Nov. 23, 2-I-3-2, Fukuoka, Japan.
2. Reddy, HM., Sharma, A., Dehzangi, A., **Shigemizu, D.**, Chandra, AA. and Tsunoda, T: GlyStruct: Glycation prediction using structural properties of amino acid residues. **International Conference on Bioinformatics (INCOB), 2018**, Sep. 26, SFIB-5, New Delhi, INDIA.
3. Asanomi, Y., **Shigemizu, D.**, Mitsumori, R., Mori, T., Niida, S., and Ozaki, K.: Association study of novel risk rare variants of late-onset Alzheimer's disease. **The Japan Society of Human Genetics, 2018**, Oct. 12, BO-04, Yokohama, Japan.
4. **Shigemizu, D.**, Akiyama, S., Asanomi, Y., Boroevich, KA., Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., and Niida, S.: The construction of risk prediction models for dementia with supervised principal component analysis using miRNA expression data. **The Japan Society of Human Genetics, 2018**, Oct. 12, O-102, Yokohama, Japan.
5. **Shigemizu, D.:** Clinical study using Next generation sequencing. **Japanese Society of Bioinformatics, Kagoshima, 2018**, May 25.

(ポスター発表)

1. **Shigemizu, D.**, Akiyama, S., Asanomi, Y., Boroevich, KA., Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., and Niida, S.: The construction of risk prediction models for dementia with supervised principal component analysis using miRNA expression data. **第6回JMACシンポジウム, 2019**, Jan. 24-25, Tokyo, Japan.
2. Asanomi, Y., **Shigemizu, D.**, Kayano, M., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., and Niida, S. Exploration of serum microRNA biomarkers for dementia-risk prediction. **第6回JMACシンポジウム, 2019**, Jan. 24-25, Tokyo, Japan.
3. **Shigemizu, D.**, Akiyama, S., Asanomi, Y., Boroevich, KA., Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., and Niida, S.: The construction of risk prediction models for dementia with supervised principal component analysis using miRNA expression data. **The American Society of Human Genetics, 2018**, P2543F, San Diego, CA, USA.

4. 4. Mori, T., **Shigemizu, D.**, Akiyama, S., Mitsumori, R., Asanomi, Y., Niida, S., and Ozaki, K.: Expression quantitative trait loci analysis with blood cells from late-onset Alzheimer's disease patients in Japanese. *The Japan Society of Human Genetics*, 2018, Oct. 12, P-260, Yokohama, Japan.
5. Mitsumori, R., Asanomi, Y., **Shigemizu, D.**, Mori, T., Akiyama, S., Niida, S., and Ozaki, K.: Genome wide association study for four types of dementia in Japanese population. *The Japan Society of Human Genetics*, 2018, Oct. 12, P-167, Yokohama, Japan.
6. Wong, JH., Akiyama, S., Nakagawa, H., **Shigemizu, D.**, and Fujimoto, A.: Identification and characterization of intermediate-sized deletions within a Japanese population. *The Japan Society of Human Genetics*, 2018, Oct. 11, P-120, Yokohama, Japan.
7. Miya, F., **Shigemizu, D.**, Kanemura, Y., Saitoh, S., Okamoto, N., Kato, M., Matsunaga, T., Mutai, H., Kosaki, K., and Tsunoda, T.: Beyond the limitations of exome analysis for genetic disorders. *The Japan Society of Human Genetics*, 2018, Oct. 11, P-118, Yokohama, Japan.
8. Asanomi, Y., **Shigemizu, D.**, Kayano, M., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Ozaki, K., and Niida, S.: Exploration of serum microRNA biomarkers for dementia risk prediction. *第10回日本RNAi研究会/第5回日本細胞外小胞学会 JSEV*, Hiroshima, 2018, Aug 30, P-64.
9. **Shigemizu, D.**, Akiyama, S., Asanomi, Y., Boroevich, KA., Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., and Niida, S.: The construction of risk prediction models for dementia with supervised principal component analysis using miRNA expression data. *第10回日本RNAi研究会/第5回日本細胞外小胞学会 JSEV*, Hiroshima, 2018, Aug 30, P-60.
10. 1 0 . Okamoto, K., **Shigemizu, D.**, Okano, T., Yeh, TW., Takashima, T., Yamashita, M., Tanita, K., Ono, S., Mitsuiki, N., Naruto, T., Okada, S., Takagi, M., Mori, M., Kanegane, H., Tsunoda, T., Imai, K., and Morio, T.: Whole exome sequence analysis using the known and candidate genes for primary immunodeficiency diseases. *The 2nd Scientific Congress of Asia-Pacific Society for Immunodeficiency*, 2018, May 7, Chongqing, China.
11. 1 1 . Yagihara, N., Watanabe, Y., Makita, N., Horie, M., Simizu, W., Ono, S., Hasegawa, K., Aiba, T., Tanaka, T., Tsunoda, T., **Shigemizu, D.**, and Minamino, T.: Identification of Mutations in Causative Genes for Cardiomyopathies in Patients with Arrhythmia Syndromes and Structurally Normal Heart. *The 82th Annual Scientific Meeting of the Japanese Circulation Society*, 2018, Mar. 24, Osaka, Japan.

G. 知的財産権の出願・登録状況

1. 特許取得

認知症の検出のためのキット又はデバイス及び方法 (2018).

出願番号: 特願 2018-085652, 出願日: 平成 30 年 4 月 26 日,

出願人: 重水 大智, 新飯田 俊平, 櫻井 孝, 鈴木 佳奈, 河内 淳平, 須藤 裕子, 吉本 真紀子, 小園 聡子, 近藤 哲司.

2. 実用新案登録

なし

3. その他

なし