

長寿医療研究開発費 2019年度 総括研究報告

「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の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、臨床情報や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や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。本年は、研究対象疾患をアルツハイマー型認知症(AD)とし、網羅的な遺伝子発現差解析を実施した。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)

檜垣 小百合 国立長寿医療研究センター MGC データ管理部 (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疾患関連因子を網羅的に探索、同定する。2)その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て評価する。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

アルツハイマー型認知症(AD)、軽度認知障害(MCI)、認知機能正常(CN)の検体に対する遺伝子発現のデータは次世代シーケンサーを用いたRNA-sequencingにて測定した。使用した検体の内訳はADが271人、MCIが248人、CNが91人の計610名である。まず、各

疾患群の血中での細胞組成について調べた。細胞組成を大きく 12 種類に分類し、血中における細胞組成を比較した。その結果、AD と CN において好中球の組成に有意差が確認された(図 1)。この差は AD、MCI、CN の 3 群においても同様な傾向が確認され、単調増加性が示された (図 2a,  $P=0.001$ )。この結果は血液検査の大規模症例 ( $n>3,000$ ) においても確認された (図 2b)。この結果から、血中での好中球の組成は疾患発症予測におけるバイオマーカーになりうることを示唆された。

AD と CN の RNA-seq のデータから遺伝子発現差解析を実施した。846 遺伝子において統計学的に有意差が確認され ( $FDR<0.05, FC>1.2$ )、480 遺伝子が AD で up-regulate, 366 遺伝子が down-regulate であった (図 3a)。この 846 遺伝子から、より機能的なモジュールを見つけるため、ネットワーク解析を実施した。ノードの大きさはエッジの数に比例し、大きいノードほど重要な因子と考えられる (図 4)。つまり、そのノードの大きさからバ

イオマーカー候補のランキング化が行える (図 4b)。

PPI ネットワーク解析でランキング化された上位の遺伝子は疾患発症予測モデルにおけるバイオマーカーになりうる。そこで上述で同定した好中球の割合、性別、年齢に加え、上位の遺伝子セットを用いて疾患発症予測モデルの構築を実施した。データを学習セットとテストセットに分割し、学習セットでモデル構築後、テス

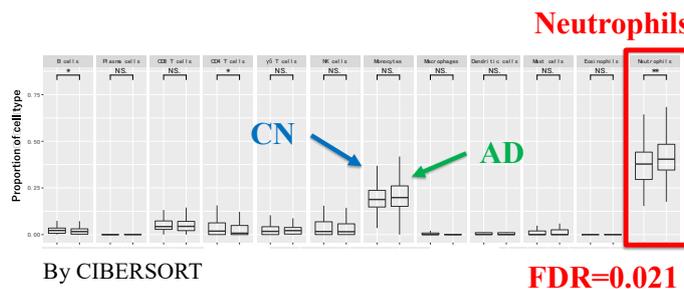


図1. AD, CNにおける血中の細胞組成の分布

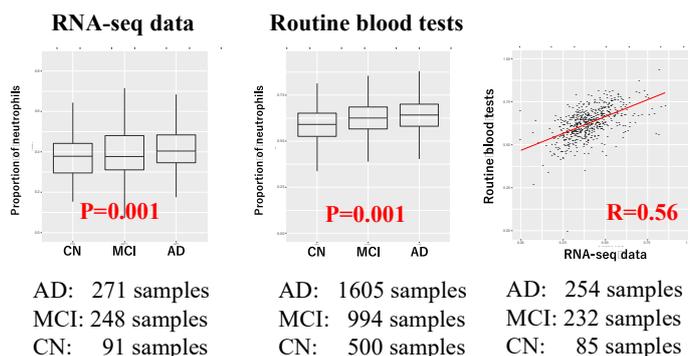


図2. AD, MCI, CNにおける血中の好中球の分布

● STAR >> featureCount >> edgeR ( $FDR < 0.05$  and  $FC > 1.2$ )

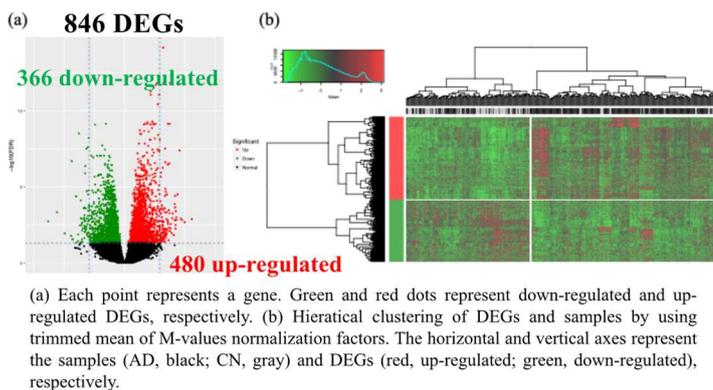


図3. AD, CNにおける遺伝子発現差解析

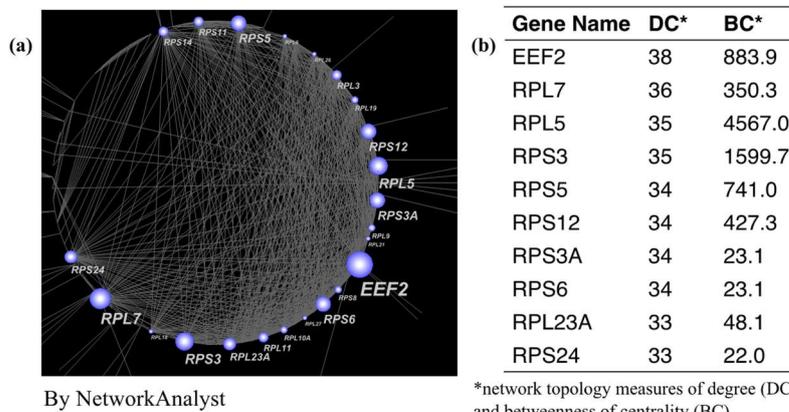


図4. DEGによるPPI-network analysis

\*network topology measures of degree (DC) and betweenness of centrality (BC).

トセットでモデルを評価した。その結果、上位 2 遺伝子(EFF2、RPL7)を採用した予測モデルが最も良い精度を示した(Acc.=0.877)。また、この予測モデルに使用した因子を前向きコホートデータに適用し、ハイリスク、ローリスクに分類した結果、2 群を正確に分類することができた (図 5, Accuracy=0.73,  $P=0.039$ )。これらの結果は Alzheimer's Research & Therapy に現在投稿中である。

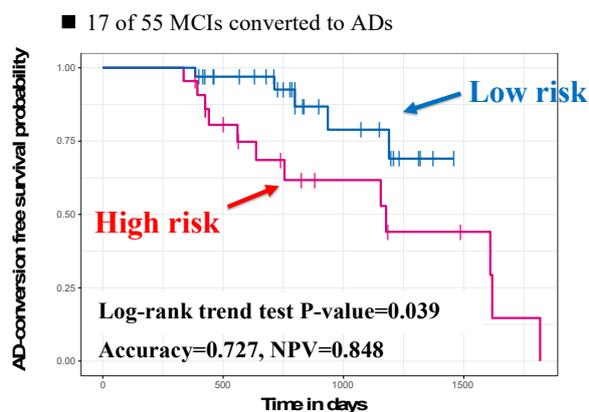


図5. 前向きコホートデータによる予測モデルの性能

#### D. 考察と結論

アルツハイマー型認知症に対して、遺伝子発現情報の網羅的な測定、遺伝子発現差解析によって疾患発症リスク因子の同定、さらにそれらを用いた高精度な疾患発症予測モデルの構築が可能になる。この予測モデルの開発によって、疾患の早期発見が可能となり、疾患の早期予防が可能になることが期待される。

#### E. 健康危険情報

なし

#### F. 研究発表

##### 1. 論文発表

- Chandra, A., Sharma, A., Dehzangi, A., **Shigemizu, D.**, and Tsunoda, T. (2019) Bigram-PGK: phosphoglycerylation prediction using the technique of bigram probabilities of position specific scoring matrix. *BMC Molecular and Cell Biology* Dec 20;20(Suppl 2):57.
- Shigemizu, D.\***, Akiyama, S., Asanomi, Y., et al. (2019) A comparison of machine learning classifiers for dementia with Lewy bodies using miRNA expression data. *BMC Med Genomics* Oct 30;12(1):150.
- Sharma, A., Vans, E., **Shigemizu, D.**, et al. (2019) DeepInsight: a methodology to transform a non-image data to an image for convolution neural network architecture. *Sci Rep* Aug 6;9:11399.
- Wong, JH., Y., **Shigemizu, D.**, Yoshii, Y., et al. (2019) Identification of intermediate-sized deletions and inference of their impact on gene expression in a human population. *Genome Medicine* 11:44.
- Asanomi, Y., **Shigemizu, D.**, Miyashita, A., et al. (2019) A functional rarevariant of SHARPIN

associated with the risk of late-onset Alzheimer's disease in Japanese. *Molecular Medicine* Dec. 25:2.

6. Singh, V., Sharma, A., Chandra, A., Dehzangi, A., **Shigemizu, D.**, et al. (2019) Computational prediction of Lysine pupylation sites in prokaryotic proteins using position specific scoring matrix into bigram for feature extraction. *Lect Notes in Artificial Intelligence* Part III, pp.488-500.
7. Vans, E., Sharma, A., Patil, A., **Shigemizu, D.**, et al. (2019) Clustering of Small-Sample Single-Cell RNA-Seq data via Feature Clustering and Selection. *Lect Notes in Artificial Intelligence* Part III, pp.445-456
8. **Shigemizu, D.\***, Akiyama, S., Asanomi, Y., et al. (2019) Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. *Commun Biol* Feb.25; 2:77.
9. Reddy, HM., Sharma, A., Dehzangi, A., **Shigemizu, D.**, et al. (2019) GlyStruct: Glycation prediction using structural properties of amino acid residues. *BMC Bioinformatics* 19(Suppl 13):547.
10. Yamaguchi-Kabata, Y., Morihara, T., Ohara, T., Ninomiya, T., Takahashi, A., Akatsu, H., Hashizume, Y., Hayashi, N., **Shigemizu, D.**, et al. (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.
11. López, Y., Kamola, PJ., Sharma, R., **Shigemizu, D.**, et al. (2018) Computational pipelines and workflows in Bioinformatics. *Encyclopedia of Bioinformatics and Computational Biology* Sep 1.
12. Nishino, J., Kochi, Y., **Shigemizu, D.**, et al. (2018) Empirical Bayes estimation of semi-parametric hierarchical mixture models for unbiased characterization of polygenic disease architectures. *Frontiers in Genetics* Apr 24, 9:115.
13. **Shigemizu, D.\***, Miya, F., Akiyama, S., et al. (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

1. **重水 大智**: 「ヒトゲノム・オミクスデータを用いた認知症研究の現在」ゲノム医科学とバイオインフォマティクスの接点と集学研究, 三島, Feb 14, 2020.
2. **重水 大智**: 「ヒトゲノム・オミクスデータを用いた認知症研究の現在」第8回生命医薬情報学連合大会(IIBMP2019), 東京, Sep 10, 2019.
3. **重水 大智**: 「ヒトゲノム・オミクスデータを用いた認知症研究の現在」Aging Forum, 大阪, July 13, 2019.

4. **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.
5. **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.
6. **重水 大智**: 「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(IIBMP2018), 鶴岡, Sep 20, 2018.
7. **Shigemizu, D.:** 「AI-Aided Genomic Studies for Disease Prediction」2018 AI Future: Application on Geriatrics Care Symposium, Taipei, July 23, 20

G. 知的財産権の出願・登録状況

1. 特許取得

1. 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