



2020年8月27日

東北大学大学院医学系研究科
東北大学東北メディカル・メガバンク機構

自閉スペクトラム症が異種の疾患の集合体である 可能性を発見

～人工知能を用いて自閉スペクトラム症の個別化医療を実現へ～

【発表のポイント】

- ▶ 人工知能の一つである機械学習*1の手法を活用し、自閉スペクトラム症（ASD）*2が異種の疾患の集合体である可能性があることを世界で初めて発見しました。
- ▶ ASDを分割して、「クラスターごとの患者群」と対照群で遺伝情報上の比較を行ったところ、「患者群全員」と対照群の比較では発見できなかった多くの有意な「違い」つまり疾患原因の候補が観察されました。
- ▶ 機械学習を用いて症例をより均質な集団にクラスタリング*3することにより、それぞれの集団の特徴に応じた個人毎のアプローチ、個別化医療が可能となることが期待されます。
- ▶ ASDの遺伝的構造と病因を解明し、ASDの精密医療の開発を促進する手がかりを提供する成果です。

【概要】

自閉スペクトラム症（ASD）は、症状などの表現型*4の点からも遺伝的要因の点からも極めて多様なものであることが指摘されてきました。ASDの主な特徴は、常同行動*5とコミュニケーション障害ですが、ASDには音への過敏や統合運動障害など他にも多くの症状を示す場合があります。遺伝的要因に関しては、現在のところ1,000を超える候補遺伝子が報告されていますが、ASDのリスク増加を十分に説明する遺伝的変異は特定されていません。

東北大学大学院医学系研究科の栗山進一教授を中心とする研究グループは、表現型の変数を適切に組み合わせるとASD患者をグループ化（クラスタリング）すると、遺伝的感受性因子を特定するチャンスが増える可能性があることに着目し、人工知能の一つである機械学習の手法を活用して、ASDが異種の疾患の集合体である可能性があることを世界で初めて発見しました。クラスター分析とゲノムワイド関連解析（GWAS）*6の組み合わせアプローチはわたしたちの知る限りでは、ASD等の疾患に適用した初めての試みです。

この研究成果は、Translational Psychiatry 誌に 2020 年 8 月 17 日にオンライン掲載されました。

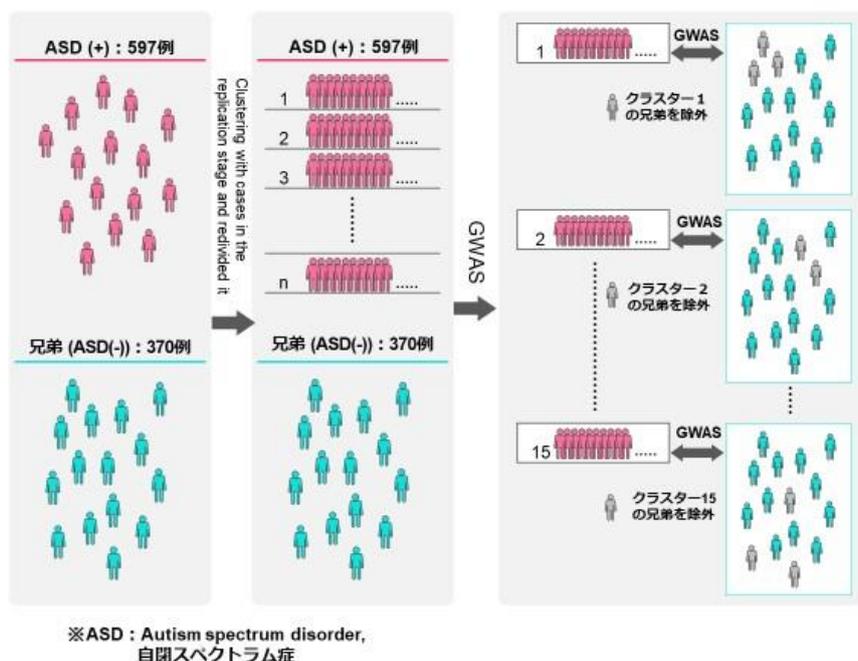
【背景】

これまで、ある疾患名でまとめられ潜在的に異質な集団を一塊として扱った「患者群全員」と対照群との GWAS は多くの疾患で実施されてきましたが、ASD については有意な SNP^{*7} はみられるものの、その効果は小さく、遺伝的感受性因子を特定したとはいえない状況であると指摘されてきました。しかし、遺伝的要因に関しては、一卵性双生児の ASD の一致率は 92%、二卵性双生児は 10% であることや、兄弟間の ASD の再発リスク比は 22 と高いことなどから、リスクの増加には遺伝的要因が強く寄与していることが示唆されています。

近年ではデータセットに内在するパターンを明らかにできる人工知能技術である機械学習手法が開発されており、機械学習のクラスタリングアルゴリズムとそれに続く GWAS は、新規でより遺伝的に均質な集団、クラスターを明らかにするのではないかという仮説を立てました。

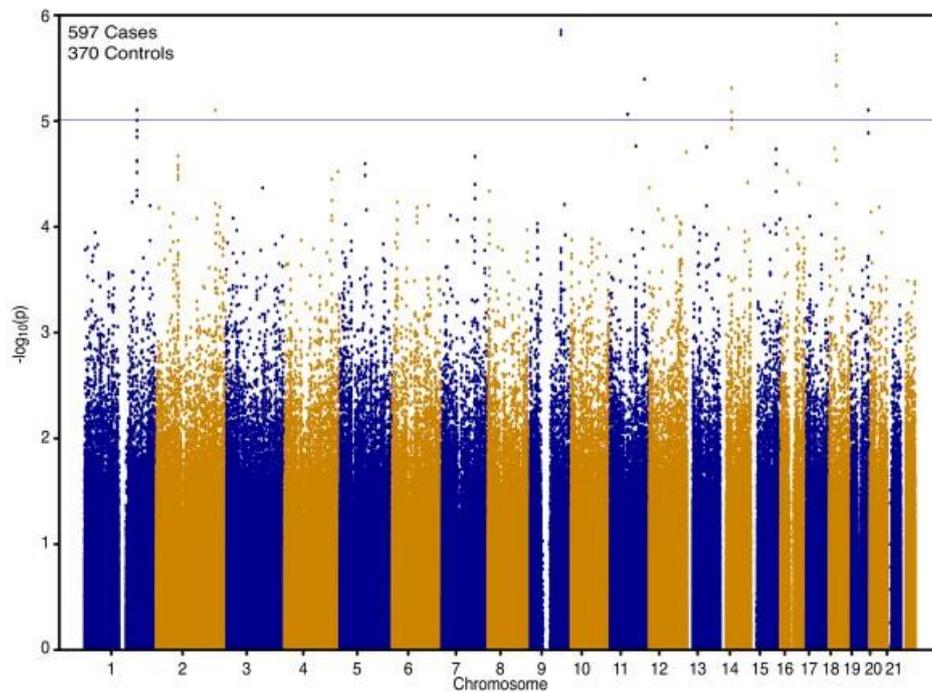
【研究の詳細】

米国のサイモンズ財団 (Simons Simplex Collection) からのデータの分譲を受け、クラスター数 15 の k-means^{*8} アルゴリズムを使用し、ASD が示す多様な症状とビタミン B6 処方歴を変数としてクラスター分析を行いました。予備研究として、597 例の ASD 症例と 370 例の対照のデータセットを使用して、従来の GWAS を実施しました。2 番目のステップでは、クラスタリングの結果に基づいてケースを分割し、各サブグループとコントロール (クラスターベース GWAS) で GWAS を実施しました。

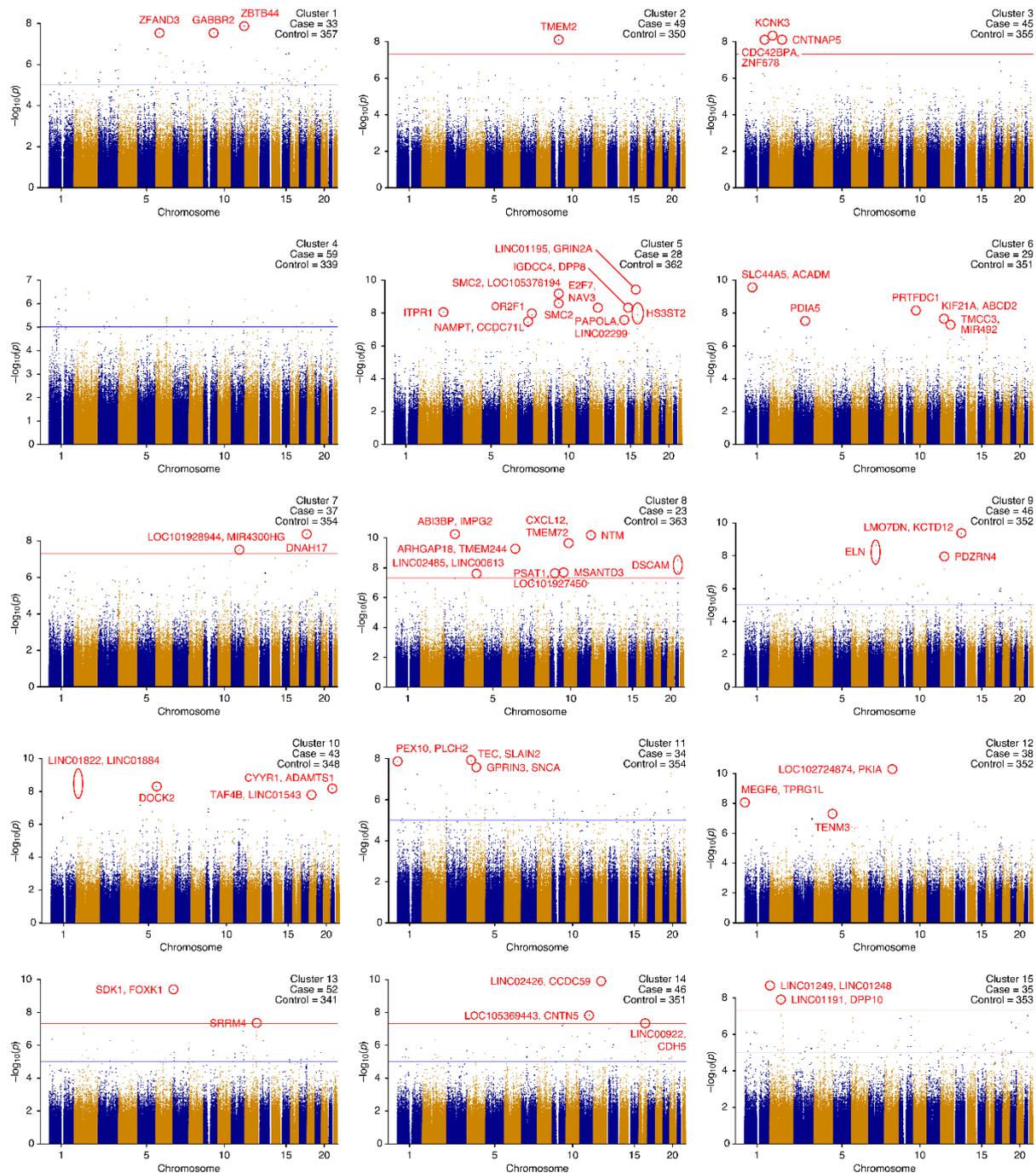


[図 1 クラスターごとの患者群と対照群とで実施された GWAS の方法]

従来の ASD という疾患名でまとめられた潜在的に異質な集団を一塊として扱った「患者群全員」と対照群を用いた GWAS 手法で行われた予備解析では、有意な関連は観察されませんでした。2 番目のクラスターに基づいた GWAS (クラスターベース GWAS) では、65 個の有意な遺伝子座を特定しました。これらの 65 の遺伝子座のうち、SRRM4 遺伝子内に位置する rs11064685 は、再現性を検証する解析でも症例と対照で有意に異なる分布を示し、ASD の原因のひとつである可能性があります。



[図 2 患者群全員と対照群とで実施された GWAS のマンハッタンプロット*9]

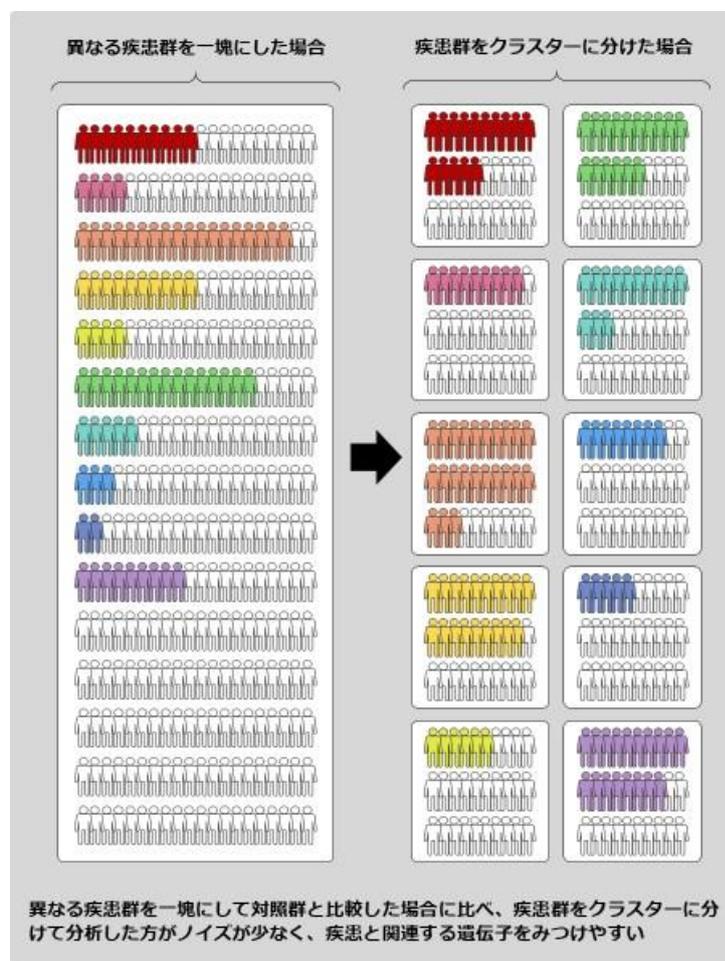


[図3 クラスターごとの患者群と対照群とで実施されたGWASのマンハッタンプロット]

【研究の意義】

これらの分析結果は、クラスタリングが比較的均一な病因をもつサブグループを首尾よく特定できる可能性を示唆しています。わたしたちの研究は、データセットが複数の異種のサブグループで構成されている場合、均質な個体の数をはるかに少ないサブグループでさえ、影響が大きく統計学的に有意な遺伝的要因を検出できることを示

しました。



[図 4 仮想上の患者群全員とクラスターごとの患者群の比較：クラスターに分けた方がノイズが少ない]

これまでの多くの GWAS 研究は、サンプルサイズを大きくすればするほど多くの有意な遺伝子座を特定できることを示してきましたが、得られた結果ではオッズ比がせいぜい 1.5 程度と低く、創薬のターゲットとするなど疾患対策を立てるためには不十分な発見であることが指摘されてきました。

わたしたちのデータは、広く定義された疾患名よりも ASD のより均一なサブグループを識別する手段としてのクラスターベース GWAS の重要性を示しています。本研究結果は、ASD の遺伝的構造と病因を解明し、ASD の精密医療の開発を促進する手がかりを提供します。

【展望】

本研究結果から機械学習を用い症例をより均質な集団にクラスタリングすることにより、それぞれの集団の特徴に応じた ASD の個別化医療が可能となることが期待

されますが、多因子疾患によってはひとつの疾患概念でまとめられていたものが実は多くのサブグループからなり、それぞれ原因や治療効果が異なっている可能性もあります。多因子疾患は実は「多病気疾患」であるのかも知れません。

機械学習の手法を用いてその疾患の特徴を表していると思われる表現型に関する変数を使いサブグループを見つけ出すことで、多くの疾患の個別化医療が大いに進展することが期待されます。

【論文情報】

タイトル : Clustering by phenotype and genome-wide association study in autism

著者 : Akira Narita, Masato Nagai, Satoshi Mizuno, Soichi Ogishima, Gen Tamiya, Masao Ueki, Rieko Sakurai, Satoshi Makino, Taku Obara, Mami Ishikuro, Chizuru Yamanaka, Hiroko Matsubara, Yasutaka Kuniyoshi, Keiko Murakami, Fumihiko Ueno, Aoi Noda, Tomoko Kobayashi, Mika Kobayashi, Takuma Usuzaki, Hisashi Ohseto, Atsushi Hozawa, Masahiro Kikuya, Hirohito Metoki, Shigeo Kure, Shinichi Kuriyama

掲載誌 : Translational Psychiatry

掲載日 : 2020年8月17日

DOI : 10.1038/s41398-020-00951-x

URL : <https://rdcu.be/b6jMh>

【用語説明】

- *1. 機械学習 : 機械学習 (Machine Learning) は人工知能解析技術のひとつで、人間が経験的に行っている学習するという能力と同様の機能、特にデータを学習してそこに潜む特徴を見つけ出し、タスクをより性能よく遂行するためのモデルを構築するような機能をコンピュータで実現しようとする技術のことです。
- *2. 自閉スペクトラム症 (ASD) : 自閉症と同義。発達障害のひとつで、常同行動とコミュニケーション障害の大きな2つの特徴をもちます。ASDには、これら2つの特徴以外にも、音への過敏、言語の表出障害、統合運動障害など多くの特徴をもつ場合があります。
- *3. クラスタリング : データの性質からデータのかたまり (クラスター) をつくる手法のことです。一見すると何らの関連もないようなデータの集合体の中から、ある基準を用いることで似たようなデータの集団を同定し、クラス分けを行います。
- *4. 表現型 : 生物の外見に現れた形態的あるいは生理的な性質のことです。遺伝型に対する言葉として用いられ、髪の色や体型、性格などが含まれます。
- *5. 常同行動 : 主に反復的あるいは儀式的な行動、姿勢、発声などをいいます。身体揺すりや足を重ねたり解いたり繰り返すなどが含まれます。自閉スペクトラム症ではこの常同行動が大きな特徴のひとつですが、個人ごとに行動の種類やその強さに極めて大きな違いがみられます。

- *6. ゲノムワイド関連解析 (GWAS) : Genome-Wide Association Study の略。ゲノム全体を対象に SNP 等をマーカーとして、患者と対照との間で比較する研究手法のことです。
- *7. SNP : 一塩基多型。ゲノム配列において、ある領域で DNA の塩基配列が個人間で一塩基のみ異なる多型のなかで一定以上のアレル (同じ座位上で対立して存在する塩基) 頻度で存在するもの。
- *8. k-means : 機械学習のアルゴリズムのひとつで、まずデータを適当なクラスターに分けた後、クラスターの平均を用いてより最適なデータセットに分けていくアルゴリズムです。何個のクラスターに分けるのかを解析前に決めておく必要があります。
- *9. マンハットプロット : 全ゲノム関連解析 (GWAS) において、すべての検定結果を視覚的にみやすくするために描かれる図のことです。横軸に染色体上の位置をとり、縦軸に p 値の log 対数をとってマイナスをかけたものをとります。こうすることで、p 値が有意な SNP の存在の有無とおよその位置がわかります。

【お問い合わせ先】

(研究に関すること)

東北大学大学院医学系研究科
東北大学東北メディカル・メガバンク機構
教授 栗山 進一 (くりやま しんいち)
電話番号 : 022-717-8104
E メール : kuriyama@med.tohoku.ac.jp

(報道担当)

東北大学東北メディカル・メガバンク機構
長神 風二 (ながみ ふうじ)
電話番号 : 022-717-7908
ファクス : 022-717-7923
E メール : pr@megabank.tohoku.ac.jp