SCIENTIFIC REPORTS

OPEN

Potential identification of vitamin B6 responsiveness in autism spectrum disorder utilizing phenotype variables and machine learning methods

Taku Obara[1,2,3], Mami Ishikuro[1,2], Gen Tamiya[1,4], Masao Ueki[1,4], Chizuru Yamanaka[1,2], Satoshi Mizuno[1,2], Masahiro Kikuya[1,5], Hirohito Metoki[1,6], Hiroko Matsubara[1,2], Masato Nagai[1,2], Tomoko Kobayashi[1,7], Machiko Kamiyama[8], Mikako Watanabe[9], Kazuhiko Kakuta[10], Minami Ouchi[11,12], Aki Kurihara[13], Naru Fukuchi[14,15], Akihiro Yasuhara[16], Masumi Inagaki[17], Makiko Kaga[17,18], Shigeo Kure[1,7] & Shinichi Kuriyama[1,2,19]

We investigated whether machine learning methods could potentially identify a subgroup of persons with autism spectrum disorder (ASD) who show vitamin B6 responsiveness by selected phenotype variables. We analyzed the existing data from our intervention study with 17 persons. First, we focused on signs and biomarkers that have been identified as candidates for vitamin B6 responsiveness indicators. Second, we conducted hypothesis testing among these selected variables and their combinations. Finally, we further investigated the results by conducting cluster analyses with two different algorithms, affinity propagation and k-medoids. Statistically significant variables for vitamin B6 responsiveness, including combination of hypersensitivity to sound and clumsiness, and plasma glutamine level, were included. As an a priori variable, the Pervasive Developmental Disorders Autism Society Japan Rating Scale (PARS) scores was also included. The affinity propagation analysis showed good classification of three potential vitamin B6-responsive persons with ASD. The k-medoids analysis also showed good classification. To our knowledge, this is the first study to attempt to identify subgroup of persons with ASD who show specific treatment responsiveness using selected phenotype variables. We applied machine learning methods to further investigate these variables' ability to identify this subgroup of ASD, even when only a small sample size was available.

[1]Tohoku Medical Megabank Organization (ToMMo), Tohoku University, Sendai, Miyagi, Japan. [2]Department of Molecular Epidemiology, Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan. [3]Department of Pharmaceutical Sciences, Tohoku University Hospital, Sendai, Miyagi, Japan. [4]Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Chuo-ku, Tokyo, Japan. [5]Department of Hygiene and Public Health, School of Medicine, Teikyo University, Tokyo, Japan. [6]Division of Public Health, Hygiene and Epidemiology, Tohoku Medical and Pharmaceutical University, Sendai, Miyagi, Japan. [7]Department of Pediatrics, Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan. [8]Department of Education, Art and Science, Yamagata University, Yamagata, Yamagata, Japan. [9]Department of Pediatrics, Saka General Hospital, Shiogama, Miyagi, Japan. [10]Kakuta Child & Allergy Clinic, Tagajo, Miyagi, Japan. [11]Department of Pediatrics, NTT Medical Center Tokyo, Shinagawa-ku, Tokyo, Japan. [12]Bunkyo Education Center, Bunkyo-ku, Tokyo, Japan. [13]Fujimoto Shinjuku Hospital, Shinjuku-ku, Tokyo, Japan. [14]Department of Psychiatry, Miyagi Psychiatric Center, Natori, Miyagi, Japan. [15]Miyagi Disaster Mental Health Care Center, Sendai, Miyagi, Japan. [16]Yasuhara Children's Clinic, Neyagawa, Osaka, Japan. [17]Department of Developmental Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan. [18]Tokyo Metropolitan Tobu Medical Center for Children with Developmental Disabilities, Koto-ku, Tokyo, Japan. [19]Department of Disaster Public Health, International Research Institute of Disaster Science, Tohoku University, Sendai, Miyagi, Japan. Taku Obara and Mami Ishikuro contributed equally. Correspondence and requests for materials should be addressed to S. Kuriyama (email: kuriyama@med.tohoku.ac.jp)

Boys/Girls (No.)	13/4
Age (years) (SD)	8.8 (4.1), range; 5–19
Body weight (kg) (SD)	32.8 (15.6), range; 15.6–62.0
Body height (cm) (SD)	130.6 (21.8), range; 98.2–170.0
Signs (No.)	
Hypersensitivity to sound	8/17
Expressive verbal disorders	16/17
Clumsiness	10/17

Table 1. Characteristics and selected signs of participants before intervention. No., number. SD, standard deviation.

Autism spectrum disorder (ASD) is a disorder characterized by difficulties in social interaction and communication, and repetitive behaviours[1]. In recent years, ASD has been investigated using advanced technologies such as machine learning to improve diagnosis and prognosis prediction[2,3]. Machine learning employs artificial intelligence techniques to discover useful masked patterns.

The effects of vitamin B6 in ASD are controversial. Previous attempts to evaluate the efficacy of vitamin B6 in treating persons with ASD have failed to produce consistent findings[4–8]. ASD is highly heterogeneous[9], and identifying a subgroup of persons who may show responsiveness to vitamin B6, if any, is important.

We previously identified similarities between ASD and pyridoxine-dependent epilepsy (PDE)[10], which are generally considered quite different. Pyridoxine is one form of vitamin B6. Many of the signs of PDE parallel those of ASD. These signs include epilepsy, autistic tendency, hypersensitivity to sound, expressive verbal disorders, and clumsiness[1,11,12]. Therefore, we hypothesized that some persons with ASD share, even in part, some etiologies with persons with PDE and that this subgroup of persons with ASD also responds to vitamin B6 intervention. We previously conducted a preliminary randomized controlled trial to evaluate the efficacy of a high-dose vitamin B6 treatment in persons with ASD who exhibited hypersensitivity to sound, expressive verbal disorders, and clumsiness, and we found that the intervention improved verbal IQ[10,13]. Our preliminary results indicate that hypersensitivity to sound, expressive verbal disorders, and clumsiness might predict responsiveness to vitamin B6.

In addition to signs, biomarkers might also be candidates contributing to the classification of ASD. Several previous studies have consistently suggested the presence of increased plasma amino acids of glutamate and decreased glutamine among persons with some ASD[14–16]. These findings suggest that plasma glutamate and glutamine levels might contribute to the classification of ASD in addition to signs.

In the present study, we explored whether the signs and biomarkers described above can be used to classify vitamin B6-responsive persons with ASD from nonresponsive ASD persons. We apply machine learning methods of affinity propagation (AP)[17] and k-medoids[18] to further investigate these variables' ability to identify a subgroup of persons with ASD.

Results

Characteristics of participants.
The participants' pre-intervention characteristics are shown in Table 1. The 17 participants included 13 boys and 4 girls. The mean age was 8.8 years (standard deviation, 4.1 years). Sixteen of 17 participants showed expressive verbal disorders.

Hypothesis testing.
Three participants were classified as vitamin B6 possible responders, whereas 14 were classified as less responders. All three possible responders showed coexisting hypersensitivity to sound and clumsiness, compared with only two of the 14 less responders, representing a statistically significant difference ($P = 0.01$) (Table 2).

A comparison of plasma amino acids levels between the possible responders and less responders showed large differences in glutamine levels (reference range: 420–700 nmol/mL). The mean glutamine levels measured before the vitamin B6 treatment were 400.5 nmol/mL (standard error (SE), 11.9 nmol/mL) in the possible responders and 481.4 nmol/mL (SE, 10.4 nmol/mL) in the less responders; this difference was significant ($P = 0.004$) (Table 3).

Cluster analysis.
Using the presence of hypersensitivity to sound concomitant with the presence of clumsiness, plasma glutamine levels, and the Pervasive Developmental Disorders Autism Society Japan Rating Scale (PARS)[19] scores, the AP analysis showed good classification of potential vitamin B6-responsive persons with ASD (cluster 1). All the participants were relatively well classified into five groups, and clusters 2 to 5 consisted of persons who exhibited a low response to vitamin B6. A graphical representation of the AP clustering results using principal component analysis (PCA)[20] is presented in Fig. 1a. The mean glutamine levels before the vitamin B6 treatment were 495.6 nmol/mL in cluster 2, 476.3 nmol/mL in cluster 3, 417.8 nmol/mL in cluster 4, and 501.2 nmol/mL in cluster 5, respectively.

The k-medoids analysis also showed good classification. The selected number of clusters by k-medoids was also five and the result was identical to that by AP except for one participants who was classified into cluster 2 by AP and into cluster 3 by the k-medoids method. A graphical representation of the k-medoids clustering results using PCA is presented in Fig. 1b.

| Signs | Potential vitamin B6 responsiveness[a] | | P-value[b] |
	"possible responders"	"less responders"	
Hypersensitivity to sound			
Yes	3	5	0.08
No	0	9	
Clumsiness			
Yes	3	7	0.18
No	0	7	
Hypersensitivity to sound and clumsiness			
Yes	3	2	0.01
No	0	12	

Table 2. Number of participants according to signs and potential vitamin B6 responsiveness. [a]According to the Clinical Global Impression-Improvement scale. [b]Fisher's exact test.

Variables	Potential vitamin B6 responsiveness[a]	Plasma levels (SE)	P-value[b]
Glutamate (nmol/mL)	"possible responders"	21.5 (3.0)	0.86
	"less responders"	20.5 (2.4)	
Glutamine (nmol/mL)	"possible responders"	400.5 (11.9)	0.004
	"less responders"	481.4 (10.4)	

Table 3. Differences in selected plasma amino acids levels according to potential vitamin B6 responsiveness. SE, standard error. [a]According to the Clinical Global Impression-Improvement scale. [b]Student t-test.

Discussion

Using the machine learning clustering algorithms of AP and k-medoids, we successfully classified an ASD subgroup that responded to vitamin B6 from others by selected phenotype variables. The additional k-medoids analysis reflected the robustness of our clustering result because the k-medoids method is completely different from AP. Our data indicate that common variables used in previous studies may be able to identify subgroup that exhibit responsiveness to specific treatment.

Our sample size is relatively small. Because we selected candidate phenotype variables according to evidence from previous studies, they are free from the multiplicity of hypothesis tests. Furthermore, contrary to hypothesis testing, cluster analysis does not necessarily require a large sample size because it separates data according to structures. In cluster analyses, distance between clusters is more important for stable clustering than sample sizes.

Variable selection is critical issues in clustering analysis. In this study, we focused on signs and biomarkers that have been identified in previous studies as candidates and conducted hypothesis testing among these selected variables and their combinations to further select relevant variables and to reduce the dimension. Recent medical studies have frequently used the method of reducing dimension and subsequent clustering. Based on the clinical and/or statistical approach, researchers identify subsets or patterns of variables to address their study aims[21,22]. We believe that our protocol is one appropriate strategy to identify subgroup of diseases including ASD.

AP is an unsupervised clustering algorithm that identifies clusters of similar points using a set of points and a set of similarity values between the points and provides a representative example, called an exemplar, for each cluster[17]. We identified five clusters from 17 persons with ASD using the AP analysis. Of the five clusters, cluster 1 consisted of all persons who responded to vitamin B6. Although we could not identify meaningful characteristics of the other four clusters (clusters 2 to 5), the AP analysis suggested that persons with ASD may be highly heterogeneous, as previously reported[9].

Cluster analysis classifies data into groups according to their structure. However, cluster analysis itself does not predict a sample as a specific cluster. Although our cluster analyses could potentially identify a subgroup of ASD, their predictive value requires further analyses utilizing different algorithms, such as support vector machine[23,24], and evaluation of the accuracy of these results. Based on our data, further studies to determine the predictive value are warranted, presumably with a larger sample size and validation data sets.

Why low levels of plasma glutamine levels predict vitamin B6 responsiveness is unclear. Glutamine plays a central role in nitrogen metabolism in many cell systems and in the central nervous system, and glutamine synthesis has a neuroprotective function because it removes ammonia and glutamate[25]. Vitamin B6 also plays a predominant role as a coenzyme for these substances[26]. Therefore, we speculate that consuming a high dose of vitamin B6 may improve the activity of the metabolic pathways described above.

The prevalence of pyridoxine responsiveness is unknown. Although three of the 17 participants exhibited a possible response to vitamin B6 in our intervention study, the study population was not randomly recruited from among persons with ASD, rather, we restrictively recruited the participants with potential responsiveness. Therefore, the proportion of vitamin B6 responders in the general population with ASD may be relatively small. Unfortunately, in daily practice, clinically used ASD assessment tools generally do not include hypersensitivity to sound or clumsiness, and clinical tests do not typically include serum or plasma glutamine levels. Therefore, more data may be necessary to assess vitamin B6 responsiveness and estimate its prevalence.
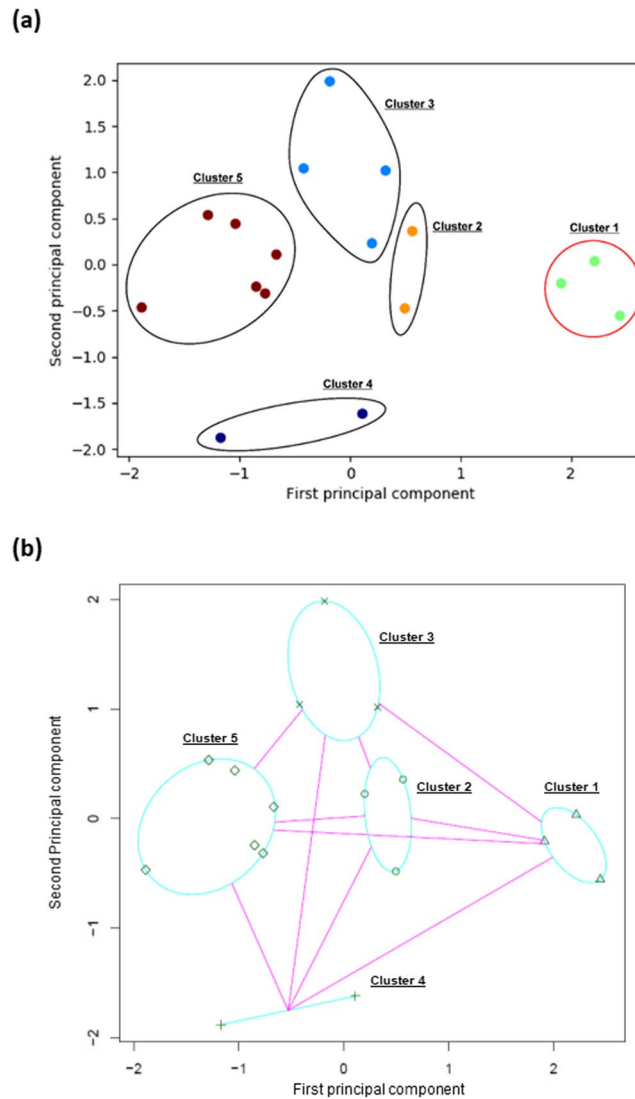
Figure 1. A graphical representation of the affinity propagation (**a**) and k-medoids (**b**) clustering results using principal component analysis. The affinity propagation (AP) analysis showed good classifying of potential vitamin B6-responsive persons with ASD (cluster 1). All the participants were relatively well classified into five groups, and clusters 2 to 5 consisted of persons who exhibited a low response to vitamin B6. A graphical representation of the AP clustering results using principal component analysis (PCA) is presented in **1a**. The k-medoids analysis also showed good classification . The selected number of clusters by k-medoids was also five and the result was identical to that by the AP except for one participants who was classified in Cluster 2 by the AP was classified in Cluster 3 by the k-medoids method. A graphical representation of the k-medoids clustering results using PCA is presented in **1b**.

We might be able to identify the genetic factors responsible for vitamin B6-responsive ASD. Although the high heritability of ASD is supported by high concordance rates (from 36% to 95%) in monozygotic twins and a higher recurrence risk of 11% and 19% with single-sibling involvement[27–29], no susceptibility genes have been specified in ASD[30]. Based on the results of the present study, the clustering of some signs and biomarkers might be informative and provide the best model for identifying etiologically similar and medically treatable predictive cases of ASD.

The possibility that a low proportion of vitamin B6 responders with ASD and the small sample size were limitations of the present study. Therefore, the present findings should be interpreted with caution, and external validity should be confirmed by further studies.

To our knowledge, this is the first study to attempt to identify subgroup of persons with ASD who show vitamin B6 responsiveness using selected phenotype variables. We applied the machine learning methods of AP and k-medoids to further investigate these variables' ability to identify a subgroup of ASD, even when only a small sample size was available.

Methods

Study design.
We analyzed the existing data from our intervention study as follows: 1) we focused on signs and biomarkers that have been identified in previous studies[10,13–16] as candidates for vitamin B6 responsiveness indicators; 2) we conducted hypothesis testing among these selected variables and their combinations to further select variables and to reduce the dimension; and 3) we conducted cluster analysis to further investigate the variables' predictive ability using two different algorithms.

Data characteristics.
Intervention study. We conducted a single-arm intervention from October 24, 2007 to September 2, 2009. Seventeen persons with ASD were recruited from seven medical institutions in Japan. Medical doctors invited potential vitamin B6 responders who showed hypersensitivity to sound, expressive verbal disorders, and/or clumsiness. The inclusion criteria for participation in this intervention study were as follows: (1) a diagnosis of ASD (code F84) determined by a doctor according to the International Classification of Diseases, 10th Revision (ICD-10)[31]; (2) 5 to 20 years of age; (3) no history of epilepsy; (4) no diagnosis of homocystinuria or fragile X syndrome; and (5) not currently receiving vitamin B6 treatments.

We selected the above age range to attempt to avoid an unstable phenotype for younger children and to exclude the effects of deuteropathy of more elderly persons.

Vitamin B6 responsiveness. Participants received 5 mg vitamin B6/kg body weight per day for two weeks followed by 10 mg vitamin B6/kg body weight per day for two weeks, resulting in a total treatment period of four weeks. Using the Clinical Global Impression-Improvement (CGI-I) scale[32], medical doctors evaluated responsiveness to vitamin B6 as "very much improved," "much improved," "minimally improved," or "no change." The CGI Scale is a measure of global clinical change with strong validity that has been widely used as an outcome measure in clinical trials of central nervous system disorders[33]. We divided participants into two groups: "possible responders," who were "very much improved" or "much improved", and "less responders," who were "minimally improved" or showed "no change".

Variable measurements. Before the intervention, participants were assessed for signs including hypersensitivity to sound, expressive verbal disorders, and clumsiness, and plasma amino acids including glutamate and glutamine. A person with ASD was considered to show a sign, such as hypersensitivity to sound, when both his/her medical doctor and guardian independently assessed the sign as present. The PARS was used to evaluate autistic traits[19].

Ethical issues. We conducted this intervention study in accordance with the guidelines of the Declaration of Helsinki[34] and all other applicable guidelines. The protocol was reviewed and approved by the institutional review board of Tohoku University Graduate School of Medicine. The board also reviewed and approved the study protocol for Miyagi Psychiatric Center, Kakuta Child & Allergy Clinic, NTT Medical Center Tokyo, Fujimoto Shinjuku Hospital, Yasuhara Children's Clinic, and National Center of Neurology and Psychiatry. For participants less than 16 years of age, we obtained informed assent from the individuals and informed consent from their guardians. For participants aged 16 to 19 years, we obtained informed consent from both the individuals and their guardians.

Hypothesis testing.
Regarding the signs' variables, we excluded expressive verbal disorders in the present analyses because they were observed in 16 of 17 participants. Signs and their combination of hypersensitivity to sound and clumsiness, and also biomarkers of plasma glutamate level, glutamine level were compared between vitamin B6 possible responders and less responders. Signs of the participants were compared using Fisher's exact test. Student's *t*-test was used to investigate the significance of the differences in the biomarkers.

The above statistical analyses were performed using SAS version 9.4 (SAS Inc., Cary, NC, USA). We used approximate variance formulas to calculate 95% confidence intervals (CIs). A P value < 0.05 was accepted as a statistically significant value. All P values were two-tailed.

Cluster analysis.
In the cluster analysis, statistically significant variables for vitamin B6 responsiveness, including hypersensitivity to sound combined with clumsiness, and plasma glutamine level, were included in the preprocessed dataset. As an a priori variable, the PARS score was also included. A relatively recently developed algorithm, AP[17], was first applied to the preprocessed dataset to divide the participants into vitamin B6 possible-responsive and less-responsive groups. AP is an unsupervised (i.e. without using the information of possible responders and less responders) clustering analysis method using a message-passing-based algorithm. In the present analysis, AP was performed without diagonal components using a dumping factor of 0.9. These analyses were performed with the scikit-learn toolkit in Python 2.6[35,36].

To examine the robustness of the clustering by AP, we additionally conducted a completely different clustering analysis using k-medoids[18] implemented in the "fpc" package for R version 3.3.2[37]. It is a fully automated method similar to AP in which the optimal number of clusters is automatically determined by the average silhouette width.

To create a graphical representation of the clustering results, we adopted a PCA[20].

Data Availability Statement
The data used in this study will be shared upon request.

References
1. Rapin, I. Current concepts: autism. *N Engl J Med.* **337**, 97–104 (1997).
2. Anzulewicz, A., Sobota, K. & Delafield-Butt, J. T. Toward the autism motor signature: gesture patterns during smart tablet gameplay identify persons with autism. *Sci Rep.* **6**, 31107, https://doi.org/10.1038/srep31107 (2016).

3. Thabtah, F. Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. *Inform Health Soc Care*. Feb 13, 1–20, https://doi.org/10.1080/17538157.2017.139913210.1080/17538157.2017.1399132 [Epub ahead of print] (2018).

4. Rimland, B., Callaway, E. & Dreyfus, P. The effect of high doses of vitamin B6 on autistic children: a double blind crossover study. *Am J Psychiatry*. **135**, 472–475 (1978).

5. Findling, R. L. *et al*. High-dose pyridoxine and magnesium administration in children with autistic disorder: an absence of salutary effects in a double-blind, placebo- controlled study. *J Autism Dev Disord*. **27**, 467–478 (1997).

6. Nye, C. & Brice, A. Combined vitamin B6-magnesium treatment in autism spectrum disorder. *Cochrane Database Syst Rev*. **4**, CD003497 (2005).

7. Kummer, A. & Harsányi, E. Comments on "vitamin supplementation reduces the level of homocysteine in the urine of autistic children". *Nutr Res*. **31**, 873 (2011).

8. Gogou, M. & Kolios, G. The effect of dietary supplements on clinical aspects of autism spectrum disorder: a systematic review of the literature. *Brain Dev*. **39**, 656–664 (2017).

9. David, M. M. *et al*. Comorbid analysis of genes associated with autism spectrum disorders reveals differential evolutionary constraints. *PLoS One*. **11**, e0157937, https://doi.org/10.1371/journal.pone.0157937 (2016).

10. Kuriyama, S. *et al*. Pyridoxine treatment in a subgroup of children with pervasive developmental disorders. *Dev Med Child Neurol*. **44**, 284–286 (2002).

11. Baxter, P., Griffiths, P., Kelly, T. & Gardner-Medwin, D. Pyridoxine-dependent seizures: demographic, clinical MRI and psychometric features, and effect of dose on intelligence quotient. *Dev Med Child Neurol*. **38**, 998–1006 (1996).

12. Burd, L., Stenehjem, A., Franceschini, L. A. & Kerbeshian, J. A 15-year follow-up of a boy with pyridoxine (vitamin B6)-dependent seizures with autism, breath holding, and severe mental retardation. *J Child Neurol*. **15**, 763–765 (2000).

13. Kamiyama, M., Kuriyama, S. & Watanabe, M. A clinical study of pyridoxine treatment for pervasive developmental disorders with hypersensitivity to sound. *No To Hattatsu*. **38**, 277–282 (2006).

14. Shimmura, C. *et al*. Alteration of plasma glutamate and glutamine levels in children with high-functioning autism. *PLoS One*. **6**, e25340, https://doi.org/10.1371/journal.pone.0025340 (2011).

15. Ghanizadeh, A. Increased glutamate and homocysteine and decreased glutamine levels in autism: a review and strategies for future studies of amino acids in autism. *Dis Markers*. **35**, 281–286, https://doi.org/10.1155/2013/536521 (2013).

16. Al-Otaish, H. *et al*. A. relationship between absolute and relative ratios of glutamate, glutamine and GABA and severity of autism spectrum disorder. *Metab Brain Dis*. Feb 3, https://doi.org/10.1007/s11011-018-0186-6 [Epub ahead of print] (2018).

17. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science*. **315**, 972–976 (2007).

18. Kaufman, L. & Rousseeuw, P. J. Finding groups in data: an introduction to cluster analysis (John Wiley & Sons, 2009).

19. Ito, H. *et al*. Validation of an interview-based rating scale developed in Japan for pervasive developmental disorders. *Res Autism Spectr Disord*. **6**, 1265–1272 (2012).

20. Cordella, C. B. Y. PCA: The basic building block of chemometrics in Analytical Chemistry. *IntechOpen*. Nov 7, https://doi.org/10.5772/51429 http://www.intechopen.com/books/analytical-chemistry/pca-the-basic-building-block-of-chemometrics (2012).

21. Terada, A., Okada-Hatakeyama, M., Tsuda, K. & Sese, J. Statistical significance of combinatorial regulations. *Proc Natl Acad Sci USA* **110**, 12996–13001, https://doi.org/10.1073/pnas.1302233110 (2013).

22. Zampetaki, A. *et al*. Prospective study on circulating MicroRNAs and risk of myocardial infarction. *J Am Coll Cardiol*. **60**, 290–299, https://doi.org/10.1016/j.jacc.2012.03.056 (2012).

23. Raudys, S. How good are support vector machines? *Neural Netw*. **13**, 17–19 (2000).

24. Bruining, H. *et al*. Behavioral signatures related to genetic disorders in autism. *Mol Autism*. **5**, 11, https://doi.org/10.1186/2040-2392-5-11 (2014).

25. Häberle, J. *et al*. Inborn error of amino acid synthesis: human glutamine synthetase deficiency. *J Inherit Metab Dis*. **29**, 352–358 (2006).

26. Sato, K. Why is vitamin B6 effective in alleviating the symptoms of autism? *Med Hypotheses*. **115**, 103–106, https://doi.org/10.1016/j.mehy.2018.04.007 (2018).

27. Rosenberg, R. E. *et al*. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch Pediatr Adolesc Med*. **163**, 907–914 (2009).

28. Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M. & Law, P. Sibling recurrence and the genetic epidemiology of autism. *Am J Psychiatry*. **167**, 1349–1356 (2010).

29. Ozonoff, S. *et al*. Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics*. **128**, e488–495, https://doi.org/10.1542/peds.2010-2825 (2011).

30. Sfari, G. Gene scoring. https://gene.sfari.org/database/gene-scoring/ (2018).

31. World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. http://www.who.int/classifications/icd/en/bluebook.pdf (1992).

32. Busner, J. & Targum, S. D. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. **4**, 28–37 (2007).

33. Neul, J. L. *et al*. Improving treatment trial outcomes for rett syndrome: the development of rett-specific anchors for the clinical global impression scale. *J Child Neurol*. **30**, 1743–1748, https://doi.org/10.1177/0883073815579707 (2015).

34. World Medical Association. WMA Declaration of Helsinki – Ethical principles for medical research involving human subjects. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (2018).

35. Pedregosa, F. *et al*. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. **12**, 2825–2830 (2011).

36. Scikit-learn. Machine learning in python (Version 0.18.1). http://scikit-learn.org/stable/# (2018).

37. Mente, S. & Kuhn, M. The use of the R language for medicinal chemistry applications. *Curr Top Med Chem*. **12**, 1957–1964 (2012).

Acknowledgements

Author Contributions

Study conception and design: T.O., M.I., M.K., M.W., M.I., M.K., S.K. and S.K. Data acquisition: M.K., M.W., K.K., M.O., A.K., N.F., A.Y., M.I., M.K. and S.K. Data analysis and interpretation: T.O., M.I., G.T., M.U., C.Y., S.M., M.K., H.M., H.M., M.N., T.K. and S.K. Drafting of the manuscript: T.O., M.I., G.T., S.M. and S.K. Critical revision of the manuscript for important intellectual content: T.O., M.I., G.T., M.U., C.Y., S.M., M.K., H.M., H.M., M.N., T.K., M.K., M.W., K.K., M.O., A.K., N.F., A.Y., M.I., M.K., S.K. and S.K. Statistical analysis: T.O., M.I., G.T., M.U., S.M. and S.K. Obtaining funding: S.K. Administrative, technical, or material support: C.Y., S.M., M.K., H.M., H.M., M.N. and T.K. Study supervision: S.K.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.