Association between Neuroligin-1 polymorphism and plasma glutamine levels in individuals with autism spectrum disorder

In-Hee Lee,[a] Douglas I. Walker,[b] Yufei Lin,[a] Matthew Ryan Smith,[c,d] Kenneth D. Mandl,[a,e,f] Dean P. Jones,[c] and Sek Won Kong[a,f,*]

[a]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, 02215, USA
[b]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, 30322, USA
[c]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, Emory University, Atlanta, GA, 30602, USA
[d]Atlanta Department of Veterans Affairs Medical Center, Decatur, GA, 30033, USA
[e]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA
[f]Department of Pediatrics, Harvard Medical School, Boston, MA, 02115, USA

eBioMedicine
2023;95: 104746
Published Online xxx
https://doi.org/10.1016/j.ebiom.2023.104746

Summary

Background Unravelling the relationships between candidate genes and autism spectrum disorder (ASD) phenotypes remains an outstanding challenge. Endophenotypes, defined as inheritable, measurable quantitative traits, might provide intermediary links between genetic risk factors and multifaceted ASD phenotypes. In this study, we sought to determine whether plasma metabolite levels could serve as endophenotypes in individuals with ASD and their family members.

Methods We employed an untargeted, high-resolution metabolomics platform to analyse 14,342 features across 1099 plasma samples. These samples were collected from probands and their family members participating in the Autism Genetic Resource Exchange (AGRE) (N = 658), compared with neurotypical individuals enrolled in the PrecisionLink Health Discovery (PLHD) program at Boston Children's Hospital (N = 441). We conducted a metabolite quantitative trait loci (mQTL) analysis using whole-genome genotyping data from each cohort in AGRE and PLHD, aiming to prioritize significant mQTL and metabolite pairs that were exclusively observed in AGRE.

Findings Within the AGRE group, we identified 54 significant associations between genotypes and metabolite levels ($P < 5.27 \times 10^{-11}$), 44 of which were not observed in the PLHD group. Plasma glutamine levels were found to be associated with variants in the *NLGN1* gene, a gene that encodes post-synaptic cell-adhesion molecules in excitatory neurons. This association was not detected in the PLHD group. Notably, a significant negative correlation between plasma glutamine and glutamate levels was observed in the AGRE group, but not in the PLHD group. Furthermore, plasma glutamine levels showed a negative correlation with the severity of restrictive and repetitive behaviours (RRB) in ASD, although no direct association was observed between RRB severity and the *NLGN1* genotype.

Interpretation Our findings suggest that plasma glutamine levels could potentially serve as an endophenotype, thus establishing a link between the genetic risk associated with *NLGN1* and the severity of RRB in ASD. This identified association could facilitate the development of novel therapeutic targets, assist in selecting specific cohorts for clinical trials, and provide insights into target symptoms for future ASD treatment strategies.

Funding This work was supported by the National Institute of Health (grant numbers: R01MH107205, U01TR002623, R24OD024622, OT2OD032720, and R01NS129188) and the PrecisionLink Biobank for Health Discovery at Boston Children's Hospital.

Keywords: Autism spectrum disorder; Endophenotype; Glutamine; Neuroligin-1; Glutamine-glutamate cycle; Excitatory-inhibitory imbalance

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition that affects 2.3% of 8-year-old children in the U.S.[1] Over the past two decades, its prevalence has been increasing in industrialized countries, partly due to earlier and more inclusive diagnosis.[2]

*Corresponding author. 401 Park Drive, LM5528.4, Boston Children's Hospital, Boston, MA, 02215, USA.
E-mail address: sekwon.kong@childrens.harvard.edu (S.W. Kong).

Research in context

Evidence before this study

Previous gene discovery studies have catalogued over 1000 genes with *de novo*, rare and common genetic variants potentially associated with ASD. However, translating these candidate genes into diagnostic and treatment biomarkers remains elusive. Endophenotypes could serve as a bridge between genetic risk factors and the intricate phenotypes of ASD, providing valuable insights into potential therapeutic targets and specific symptoms to focus on for ASD treatment.

Added value of this study

Specifically among ASD patients and their family members, plasma glutamine levels were associated with variants in the *NLGN1* gene, which encodes a postsynaptic cell-adhesion protein involved in the incorporation and retention of excitatory glutamatergic neurons. Furthermore, glutamine levels were correlated with the phenotypic scores for repetitive and restricted behaviours (RBS total score) in ASD patients. Notably, no direct association was observed between the *NLGN1* gene variants and the RBS total score.

Implications of all the available evidence

The imbalance of cortical cellular excitation to inhibition (E-I) has been suggested as a common pathophysiological mechanism of ASD. However, clinical trials targeting glutamatergic or inhibitory GABAergic receptors have yielded mixed results. The association between *NLGN1* variants and plasma glutamine levels could potentially help identify a target population for E-I imbalance-modulating drug treatment, with the severity of RBS serving as an outcome measure for treatment efficacy.

The heritability of ASD is estimated to be between 0.69 and 0.91 based on family and twin studies,[3] indicating that the condition is attributable to polygenic risks with varying contributions from *de novo* or rare variants with large effects and aggregate effects of common variants.[4] While hundreds of ASD candidate genes have been identified through gene discovery studies,[5] putative disease-associated variants have only been found in approximately 15% of cases.[6] Candidate genetic variants are often found in genes encoding synaptic cell-adhesion molecules that specify synaptic functions, such as neuroligins (NLGNs) and neurexins (NRXNs), as well as post-synaptic scaffolding proteins like PSD95 and SHANK proteins.[7,8] While protein-truncating variants in certain genes, including *SHANK3* and *CHD8*, have been observed more frequently in ASD than in cohorts with intellectual disability (ID) or other controls,[9] the precise role of these genes in ASD remains unclear. Indeed, as of our current understanding, no ASD candidate gene has accumulated enough evidence to establish its unique association with ASD.[10]

Deciphering the genetic underpinnings of ASD and their associated indicators that could potentially be linked to ASD aetiology is a crucial step in pinpointing possible drug targets. Given the higher success rate in regulatory approval for drug targets supported by human genetic evidence,[11,12] the role of gene discovery proves to be indispensable in developing effective ASD therapeutics.[13] Despite its importance, discovering "autism-specific" genes poses significant challenges for several reasons. Firstly, the diagnostic criteria and classification system for psychiatric disorders,[14] including ASD, rely on a wide-range of symptoms and signs, all of which differ in severity and clinical course.[15] Secondly, ASD phenotypes are the result of a complex interplay of genetic and environmental factors, epigenetic changes, and stochastic events occurring during foetal brain development. Thirdly, ASD is a disorder of the developing human brain, an organ composed of diverse cell types, extensive connections, and intricate interactions at the cellular and circuit levels, all of which are influenced by experiences and changes across the developmental trajectory. Lastly, the genetic heterogeneity of ASD is partly responsible for its clinical heterogeneity.[16] While direct associations between genotypes and phenotypes in ASD remain elusive and the polygenic nature of the disorder further complicates our understanding of the affected molecular pathways and brain circuits, there is a critical need for novel approaches to decipher the complex interplay between genotypes and phenotypes in ASD.

To better understand the biological underpinnings of psychiatric disorders, Gottesman and Shields introduced the concept of an endophenotype—measurable internal phenotypes that serve as a link between genotypes and psychiatric behavioural phenotypes, specifically within the context of schizophrenia genetics.[17] Various methods, such as neuroimaging, event-related potentials, eye-tracking, and small molecules, have been employed to identify endophenotypes for several psychiatric disorders.[18] For an indicator to be considered as an endophenotype, it should meet several criteria: it must be associated with illness in the population, exhibit heritability, demonstrate trait dependency, co-segregation with illness within families, and show a higher occurrence rate in unaffected family members compared to the general population. However, it's worth noting that for biomarkers in the conventional sense, only the first criterion is necessary.

ASD has been associated with several potential endophenotypes, including elevated serotonin levels,[19] decreased oxytocin levels,[20] abnormal activation of the prefrontal cortex during non-social visual attention tasks,[21] and atypical visual scanning of human faces.[22] Notably, heritable cognitive functional measures such

as intelligence quotient (IQ), non-verbal communication, and social adaptation have also been considered as potential endophenotypes. For instance, 'age at first word' (WORD), a quantitative trait, has been used as an endophenotype, revealing an association with common variants in the *CNTNAP2* gene.[23,24] Moreover, our previous study identified a correlation between common variants in the *VPS13B* gene, a causal gene for Cohen syndrome, and the total score of Stereotyped Behaviours and Restricted Interests, further highlighting the utility of endophenotypes.[25] Thus, endophenotypes present a promising avenue to connect genotypes with the complex ASD phenotype. This offers the potential to define subgroups with shared pathobiological mechanisms, clinical trajectories, and responses to treatment.

Recent advancements in high-throughput metabolomics have demonstrated substantial potential in the identification of such endophenotypes. This technology enables a comprehensive profiling of metabolites in an individual, which can provide a snapshot of the host genetic makeup and its interaction with environmental factors.[26,27] In essence, it enables to investigate how genetic variants can influence metabolite levels, thus offering mechanistic insights into the biological pathways that are perturbed in pathophysiological conditions.[28] In a recent cross-platform meta-analysis, Lotta and colleagues validated the reported association between single nucleotide variants (SNVs) and metabolite concentrations across independent cohorts using a genome-wide association study (GWAS) framework.[29]

However, the breadth of chemical space coverage has been limited to a few hundred metabolites in previous studies. By pioneering an untargeted analysis platform using a high-resolution mass spectrometry coupled with liquid chromatography (LC-HRMS),[30,31] we conducted a case-control study to identify metabolites that were significantly different in individuals with ASD compared to neurotypical controls.[32] Furthermore, in conjunction with genome-wide genotyping, this platform was employed to investigate the genetic influence on metabolite levels in generally healthy children.[33] We identified high heritability in certain metabolite species and discovered novel gene-metabolite associations. These findings underscore the need for continued metabolomics research across different age groups to enhance our understanding of gene-environment interactions and their impact on health trajectories. Therefore, untargeted metabolomics profiling, combined with genome-wide genotyping, could offer an unbiased approach for the discovery of endophenotypes in ASD. Significant focus in ASD research has been directed towards neuroactive amino acids, such as glutamate and GABA, which play crucial roles in maintaining the balance between excitatory and inhibitory neuronal activity. Deviations in this excitatory-inhibitory (E-I) balance have been thought to be associated with ASD, with alterations in the glutamate/glutamine cycle being one potential mediator of these deviations.[34–37] However, findings on the genetic underpinnings for E-I imbalance, as reported in both human and animal models, have been inconsistent.[38,39] Given these observations, the potential to elucidate these connections through comprehensive metabolomics profiling presents promising new avenues in ASD research.

In this study, we investigated the possible relationships between metabotypes (levels of various metabolites) and genotypes in individuals with ASD (mean age 8.8 years old (yo), range 1.9–28.4), their unaffected siblings (mean age 8.5 yo, range 2.4–16.2), fathers (mean age 41.5 yo, range 27.8–66.6), and mothers (mean age 39.1 yo, range 24.2–54.5). Our aim was to uncover unique genotype-metabotype associations that could indicate potential endophenotypes of ASD and their underlying genetic factors. We profiled the plasma metabolome in an untargeted manner, which yielded quantitative measurements of 14,342 features. These were then integrated with genome-wide genotyping data to identify potential genetic determinants of the metabotype unique to the ASD group and their family members, compared to the control group. To account for technical limitations due to the differences in plasma sample collection and biobanking methods, metabolite quantitative trait loci (mQTL) analysis was conducted separately in the ASD and control groups. Significant genotype-metabotype associations were then compared between the two groups to prioritize ASD-specific associations. Moreover, we examined whether the identified endophenotypes would correlate with specific ASD domain scores and measures of neurocognitive development.

Methods

Participants

The Autism Genetic Resource Exchange (AGRE) Consortium provided the collection of whole-genome sequencing (WGS) and phenotype data of families with at least one individual diagnosed with ASD through the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS).[40] Cases were defined using AGRE's "derived affected status" information marked with "Autism", "Not Quite Autism (NQA)", or "ASD." We excluded individuals with known genetic causes of ASD such as Fragile X syndrome, Trisomy 21, 15q deletion, and 22q duplication. The dataset contained 11,961 individuals with demographic and phenotypic information, and for the current study, we included 658 individuals with both WGS and available plasma samples for LC-HRMS analysis.

Generally healthy individuals who were enrolled in the PrecisionLink Health Discovery (PLHD) cohort at Boston Children's Hospital (BCH) between January 2016 and November 2019 were subjected to whole-

genome genotyping using microarray and metabolomics profiling using LC-HRMS.[33] For participants who had available plasma samples, we retrieved International Classification of Diseases (versions 9 and 10), Current Procedural Terminology (CPT4), and SNOMED CT codes from the BCH Cerner electronic health records (EHR) database as illustrated in Fig. S1. A total of 859,427 concept identifiers were extracted for 441 individuals. These included CPT4 (N = 381,709), ICD-9-CM (N = 199,837), ICD-10-CM (N = 269,851) and procedure codes. We mapped ICD codes to PheCodes to consolidate ICD-9-CM and ICD-10-CM into a clinically meaningful phenotype and disease classification.[41] Specifically, we assessed whether any participants had been diagnosed with ASD (PheCode: 313.3), intellectual disabilities (PheCode: 315.3), and other developmental disorders (PheCodes: 315, 315.1, and 315.2). The accuracy of ASD diagnosis using only ICD codes extracted from EHR has been found to be low.[42] Therefore, in addition to this, we utilized CPT4 codes: 92,523 (evaluation of speech sound production and language comprehension and expression), 92,507 (treatment of speech, language, voice, communication, and/or auditory processing disorder), 92,508 (treatment of auditory processing disorder), and 97,127 (cognitive intervention) to determine whether any participants had undergone evaluations and treatment for ASD. None of the individuals in the control cohort had a diagnosis of ASD or other neurodevelopmental disorders including attention-deficit/hyperactivity disorder (ADHD), learning disabilities, moderate to severe intellectual disabilities, or Mendelian disorders. To adhere to the Health Insurance Portability and Accountability Act regulations, personal identifiers were removed from the extracted EHR data and replaced with universal unique identifiers (UUIDs). All subsequent analyses utilized these UUIDs, age at blood collection, sex information, and sample identifiers for plasma and DNA samples provided by the BCH Biobank.

Ethics statement

For AGRE cohort, the study protocol was reviewed and approved by Autism Speaks to access the WGS data and to obtain plasma samples. All participants in AGRE provided written informed consent to Autism Speaks. The PLHD was approved by the BCH Institutional Review Board under protocol number P00000159. All participants in PLHD provided written or electronically signed informed consent, with those under 18 years old doing so through their parents or legal guardians, and those over 18 years old providing consent themselves.

Characterization of plasma metabolome using high-resolution mass spectrometry

AGRE plasma samples were thawed and aliquoted by NIMH biorepository and PLHD plasma samples were prepared at BCH Biobank. Plasma samples from the AGRE and PLHD cohorts were randomly assigned to different batches of LC-HRMS profiling to balance sample sources, age, and sex. A dual column chromatography approach involving hydrophilic interaction liquid chromatography (HILIC; XBridge BEH Amide XP HILIC column; 50 × 2.1 mm, 2.5 µm; Waters, Waltham, MA) and reversed-phase liquid chromatography (RPLC; C18 column; 50 × 2.1 mm, 2.6 µm; Higgins Analytical, Mountain View, CA) was used, and mass spectral data was collected and analysed as previously described.[33,43] The detected features were compared to an in-house library of identified metabolites.[44] The library consisted of metabolites that were confirmed by co-elution relative to authentic standards and ion dissociation mass spectrometry (level 1 identification by the criteria of Schymanski et al.[45]) with a tolerance of 5 ppm in m/z and 30 s in retention time (RT). A total of 166 features were matched with identified metabolites in the confirmed library. The details of mass spectral data collection and analysis are provided in the Supplementary Methods and the information of 166 identified metabolites are listed in Table S1.

Genome-wide genotype data

For the AGRE cohort, we downloaded variant call files (VCFs) through MSSNG[46] and the Hartwell Autism Research and Technology Initiative (iHART)[47] consortium sites. The VCFs from MSSNG (version db6, 1740 subjects) were processed with Sentieon Genomics pipeline version 201808.06 using GRCh38 as reference genome. The VCFs from iHART (version v01, 2308 subjects) were processed with GATK-3.2 using GRCh37 as reference genome. The MSSNG VCFs were lifted to GRCh37 coordinates to be merged with iHART VCFs. While merging, we excluded 374,405 variants that were genotyped inconsistently in subjects duplicated between MSSNG and iHART. A total of 681 subjects had both genotype data and LC-HRMS data and were subjected to the quality control and filtering steps: (1) variants or samples with missing rate less than 2%, (2) biallelic variants in autosome with minor allele frequency of 5% or greater, (3) variants passing Hardy-Weinberg equilibrium test ($P < 1 \times 10^{-6}$), (4) samples with heterozygosity within 3 standard deviations from average. The final genotype data for AGRE cohort consisted of 4,967,901 autosomal variants for 658 subjects.

For the PLHD cohort, genome-wide genotyping was performed using Illumina Global Diversity Arrays (GDAs), and we downloaded genotype calls from 453 subjects from the PLHD portal.[33] The genotypes were imputed with TOPMed imputation reference panel.[48] We selected autosomal variants with imputation accuracy (r^2) >0.3 and lifted variants to GRCh37 coordinates for quality control and filtering steps as in AGRE cohort. The final data for PLHD cohort consisted of 6,439,124 variants and 441 independent subjects (King robust estimator <0.177). Finally, we excluded variants specific

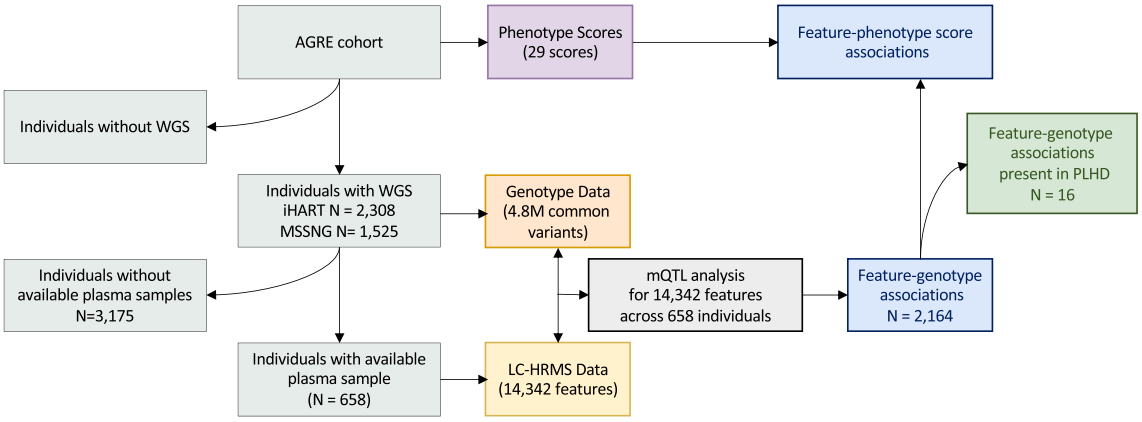

***Fig. 1:* Overview of study design and data analysis.** Diagram illustrating the workflow involving the Autism Genetic Resource Exchange (AGRE) cohort, highlighting the process of data collection and analysis for phenotype, genotype, and metabolomics information. The metabolomics data for the AGRE cohort was generated from 658 plasma samples using liquid-chromatography high-resolution mass spectrometry (LC-HRMS). These samples had accompanying whole-genome sequencing (WGS) data from either the iHART or MSSNG databases and detailed phenotype scores. The analysis includes a metabolite Quantitative Trait Loci (mQTL) analysis and a comparison with variant–metabolite associations observed in the PrecisionLink Health Discovery (PLHD) cohort. Variant-metabolite associations that passed genome-wide significance ($P < 5 \times 10^{-8}$) were selected for further evaluation in the AGRE cohort. Associations that were concurrently present in the PLHD cohort were excluded from further evaluation and correlation with phenotype scores.

to either cohort, resulting in 4,840,930 variants in both cohorts for further analysis. The detailed steps of quality control and filtering for genotype dataset are described in the Supplementary Methods and in Fig. S2.

Statistical analysis

The overall workflow, including data collection, metabolomics profiling, and association analysis, is depicted in Fig. 1. The AGRE and PLHD cohorts were analysed separately due to differences related to the plasma sample collection, preparation, and biobanking processes. Within each cohort, however, all individuals were analysed together without further stratifying by age groups or ASD status. Each feature intensity was transformed to log base 2 scale and normalized by taking residuals from a generalized linear regression fitted with subjects' age, sex, and batch of metabolomic profiling as covariates ($\log_2$ (*feature*) ~ *age* + *sex* + *batch*). For each feature in each cohort, a linear mixed model was fitted over the entire cohort using GCTA[49] (version 1.94.0 beta; –fastGWA-mlm), with the top 10 principal components (calculated using PLINK2 software[50]) as covariates to account for population structure. Finally, we performed a stepwise conditional and joint analysis using GCTA-COJO (–cojo-slct)[51] to obtain conditionally independent variants for each feature.

The polygenic risk score (PRS) for ASD was calculated using the summary statistics published by the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and the Psychiatric Genomics Consortium (PGC) in November 2017. We used PLINK (v1.90b5.1 64-bit (20 December 2017)) to clump variants and calculate PRS. We then assessed the association between the PRS and feature levels using a generalized linear model for each feature with subjects' age, sex, and 10 PCs as covariates. The association test between the PRS and feature levels were limited to the individuals with European ancestry in AGRE cohort.

Significant features associated with genotype were tested for correlation with phonotype scores, using 29 different scores from 9 instruments for core symptoms of ASD or neurocognitive development (Table S2).[25] We used a generalized linear regression model with subjects' sex and age as covariates (*phenotype score* ~ $\log_2$ (*feature*) + *sex* + *age*). For features that showed significant associations with both phenotype scores and genetic variants, we conducted a mediation analysis to test whether the metabolite mediated the genetic effect on phenotype score using PROCESS macro v4.1 for IBM SPSS version 27.[52] All statistical analysis, except for mediation analysis, were performed using the R statistical language (version 4.1.2; R Foundation for Statistical Computing, Vienna, Austria). The details of statistical analysis and software settings as well as the information on the 29 phenotype scores used in the association analysis are described in Supplementary Methods. We have applied the STREGA reporting guidelines[53] when reporting our study.

Role of funders

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of

the manuscript; and decision to submit the manuscript for publication.

Results

Study populations

The demographic characteristics of individuals from the AGRE[40] and PLHD[33] cohorts, who were selected based on the availability of stored plasma samples and genome-wide genotype data are shown in Table 1. The baseline characteristics of each cohort are compared either by a chi-squared test (for categorical variables such as sex) or by Welch's t-test (for numerical variables such as age). The AGRE cohort consisted of 350 probands including 90 females (25.7%), 17 unaffected siblings, and 291 parents. The PLHD cohort consisted of 441 individuals, 300 of whom were under 18 years old. The mean age of AGRE cohort was 22.6 years with a standard deviation (SD) of 16.53, which was significantly older than PLHD cohort ($P < 0.001$). However, the mean age of probands in AGRE cohort was 8.8 years (SD of 4.15 years), which was significantly younger than the children in the PLHD group (Welch's t-test, $P = 8.7 \times 10^{-8}$). All individuals in the PLHD cohort were generally healthy and had no history of neurodevelopmental disorders as described in Methods. There were significant differences in the proportion of self-reported race or ethnicity between AGRE and PLHD cohort; however, we found no differences in the proportion of genotype-based global ancestry ($P = 0.51$). The genotype-based global ancestry for each individual was estimated by projecting individuals into principal component space defined using the whole genomes from 1000 Genomes Project and finding the nearest population group (Supplementary Methods for further details).

The AGRE cohort generally showed higher PRS than PLHD because of the probands sub-group which showed significantly higher PRS than PLHD (Fig. 2, Welch's t-test, $P = 4.9 \times 10^{-40}$). There was no difference in PRS between non-probands in AGRE cohort (parents and unaffected siblings) and the PLHD cohort.

Genotype-metabotype associations unique to the AGRE cohort

A total of 14,342 features, including 166 identified metabolites, were subjected to mQTL analysis in each

Variable	AGRE	PLHD	P[a] (t or χ^2)
Participants (n)	658	441	
Proband	350		
Unaffected sibling	17		
Parents	291		
Children		300	
Adults		141	
Age in years (mean [sd; range])	22.6 [16.53; 1.9–66.6]	15.0 [8.47; 0.4–60.1]	<0.001 (t = 10.04)
Proband	8.8 [4.15; 1.9–28.4]		
Unaffected sibling	8.5 [3.82; 2.4–16.2]		
Parents	40.2 [6.72; 24.2–66.6]		
Children		10.8 [5.13; 0.4–17.9]	
Adults		24.0 [6.98; 18.0–60.1]	
Sex (female, n (%))	262 (39.8%)	230 (52.2%)	<0.001 (χ^2 = 15.76)
Proband	90 (25.7%)		
Unaffected sibling	7 (47.1%)		
Parents	164 (56.4%)		
Children		152 (50.7%)	
Adults		78 (55.3%)	
Genotype-based global ancestry[b] (n (%))			0.51 (χ^2 = 2.32)
European	528 (80.2%)	342 (77.6%)	
African	40 (6.1%)	36 (8.2%)	
American	85 (12.9%)	58 (13.2%)	
South Asian	5 (0.8%)	5 (1.1%)	

This table presents the basic demographic and clinical characteristics of participants in both the Autism Genetic Resource Exchange (AGRE) and PrecisionLink Health Discovery (PLHD) cohorts. The information includes, but is not limited to, age, sex, ASD status, and genotype-based global ancestry. Statistical comparisons across groups are also indicated. [a]P represents the P-value at a significance level of 5%. The χ^2-test was used for each qualitative variable, and Welch's t-test was employed for each quantitative variable to compare entire samples between the AGRE and PLHD cohorts. [b]Global ancestries based on genome-wide genotype data were determined by projecting each individual into the principal component space defined using the five population groups in the 1000 Genomes Project (additional details are provided in the Supplementary Material).

Table 1: **Baseline characteristics in AGRE and PLHD cohorts.**

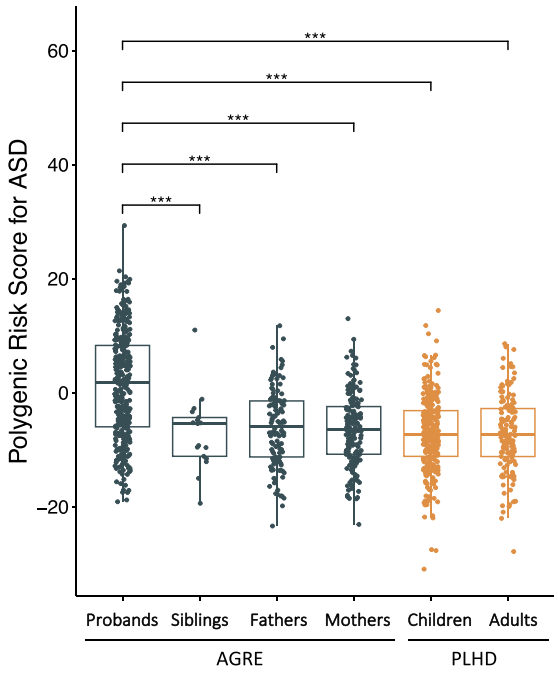

Fig. 2:* Comparison for polygenic risk scores across subgroups in the AGRE and PLHD cohorts.** The boxplots represent the distribution of the polygenic risk score (PRS) for autism spectrum disorder (ASD) across various subgroups in the AGRE and PLHD cohorts. Subgroups include probands, unaffected siblings, and parents (fathers and mothers) in the AGRE cohort, as well as children (under 18 years old) and adults in the PLHD cohort. Probands had significantly higher PRSs compared to the rest of the groups, with no differences among the other groups (P < 0.0001).

cohort. The process of filtering and prioritization of features associated with genotype in the AGRE cohort is depicted in Fig. 3. The genome-wide significant ($P < 5 \times 10^{-8}$) mQTL loci were found for 1544 features (10.8% of 14,342) and 2165 features (15.1%) in the AGRE and PLHD cohorts, respectively, with only 16 associations found in both cohorts. The shared associations corresponded to previously reported genetically influenced metabolites (GIMs), such as Ne,Ne-dimethyllysine and *PYROXD2*, bilirubin and *UGT1A1*, butyrylcarnitine and *ACADS*, and phosphocholines and *FADS1*.[29] Additionally, each cohort harboured associations mapped to the same gene or neighbouring loci as the 16 shared associations: 39 additional features in AGRE cohort and 17 in PLHD cohort. These additional associations may represent isoforms or fragments of the metabolites or small molecules in the 16 common associations. Other than the 16 associations found in both cohorts and the additional associations sharing loci with them, AGRE and PLHD cohort showed distinct mQTL results.

Next, we focused on the 139 AGRE-specific associations that involved ASD candidate genes from the SFARI Gene database (2022 Q4 release), specifically genes in categories S, 1, and 2.[5] Only one of them was associated with features identified by authentic standards: glutamine and an intronic variant on the *NLGN1* gene (NM_014932.3:c.646 + 136666T>A, rs11926085) ($P = 7.5 \times 10^{-9}$, $\beta = -0.2$ [95% CI, −0.27 to −0.13], Fig. 4a). However, despite the lack of differential frequency for rs11926085 (0.24 for AGRE and 0.28 for PLHD, $\chi^2 = 5.66$, P = 0.06) and the similarity between haplotype structures of surrounding region in AGRE and PLHD (Fig. S3), we did not observe the same association in PLHD (Fig. 4b). Individuals with the T allele had higher levels of plasma glutamine compared to those with the A allele (Fig. 4c). Also, the presence of protein-truncating variants in ASD candidate genes was not associated with subjects with plasma glutamine

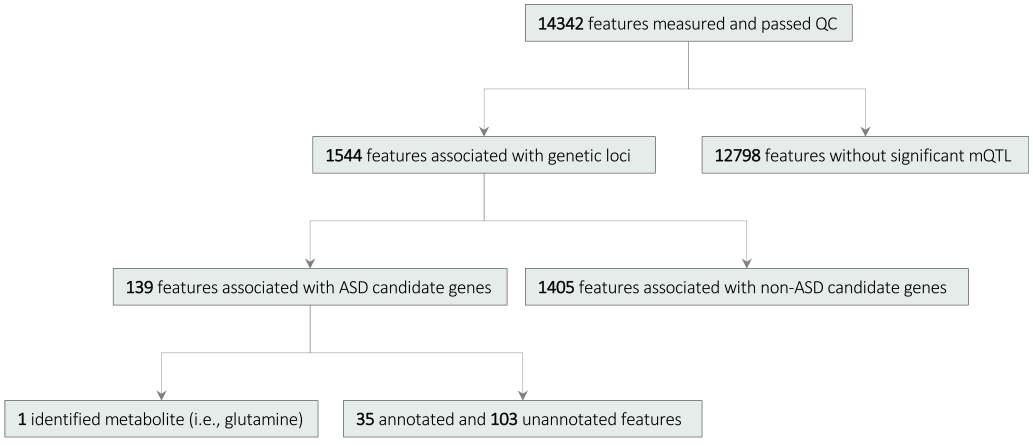

***Fig. 3:* The process of filtering and prioritization of features associated with genotype in the AGRE cohort.** This diagram depicts the systematic process used within the AGRE cohort to filter and prioritize the features that demonstrate significant genotype–metabolite associations.

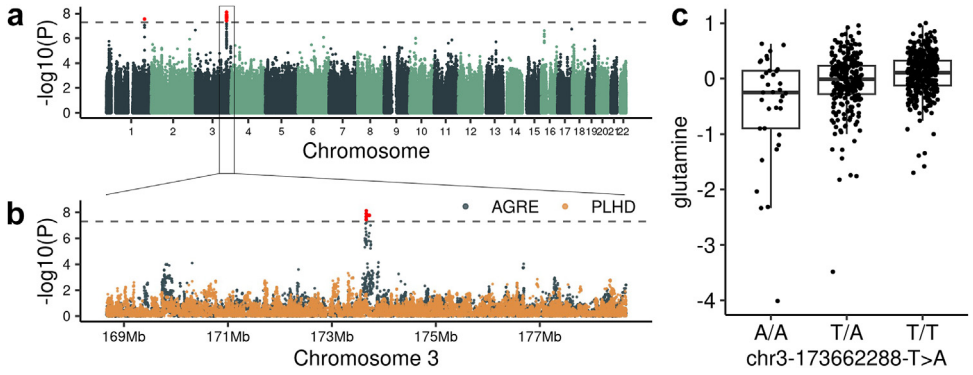

***Fig. 4:* Association of *NLGN1* loci with glutamine in the AGRE cohort.** (a) The Manhattan plot for glutamine. The dashed horizontal line represents the genome-wide significance threshold ($P = 5 \times 10^{-8}$). The y-axis shows the $-\log_{10}$ (P-values). (b) A regional plot of the *NLGN1* gene (ranging from 168.7 Mb to 178.7 Mb on chromosome 3), showing data from both the AGRE and PLHD cohorts. The dashed horizontal line denotes the genome-wide significance threshold ($P = 5 \times 10^{-8}$), while the y-axis denotes the $-\log_{10}$ (P-values). (c) Glutamine levels distribution stratified by the rs11926085 genotype (chr3:173662288 T>A). The y-axis represents glutamine levels normalized by age, sex, and batch variables.

levels outside of two SDs (Fig. S4), which suggested that the plasma glutamine levels were not driven by the genetic burden on ASD candidate genes.

The AGRE cohort also had 35 annotated and 103 unannotated features associated with ASD candidate genes, which were not found in PLHD (Table S3). Of these, 7 annotated and 2 unannotated features were found associated with the intronic variant on *NLGN1* gene, all of which were highly correlated with glutamine ($P < 0.001$). We also found one annotated (mass-to-charge ratio of 132.13 and retention time of 74.4s) and 2 unannotated features associated with intron variant on *SND1* gene (NM_014390.2:c.1779+14743A>G, rs17151653) ($P = 1.2 \times 10^{-8}$, $\beta = 0.2$ [95% CI, 0.13–0.26], Table S3). Although the 3 features were highly correlated with a feature identified as leucine, leucine itself did not reach the genome-wide significance level.

No feature was associated with ASD polygenic risk score.

Plasma glutamine levels and phenotype scores

Although glutamine is a precursor of glutamate, no locus was associated with plasma glutamate levels. Interestingly, a significant negative correlation was found in AGRE ($R = -0.51$ [95% CI, −0.555 to −0.470], $P < 0.001$, Fig. 5a), but not in PLHD ($R = 0.042$ [95% CI, −0.0457–0.129], $P = 0.35$). Among the 29 phenotype scores tested using regression modelling with plasma glutamine levels, the Repetitive Behaviour Scale total score (RBS) showed a negative correlation ($\beta = -6.53$ [95% CI, −9.73 to −3.24]; $R = -0.130$, $P = 0.043$, Fig. 5b). Considering the association between plasma glutamine level and *NLGN1* haplotype as presented above, the glutamine level may serve as an endophenotype that reflects the genetic effect of *NLGN1* on RBS whereas no association between *NLGN1* haplotype and RBS was found ($\beta = 3.4$; SE = 2.43, $P = 0.15$, Fig. 5c). Since there was no direct effect of *NLGN1* haplotype on RBS, we did not perform a mediation analysis for glutamine levels.

Discussion

In this study, we leveraged an untargeted metabolomics platform to conduct a comprehensive examination of the plasma metabolome in individuals with ASD and their family members, subsequently integrating these findings with whole-genome genotype data. We aimed to delineate plasma metabolites that could serve as endophenotypes—a measurable component unseen by the unaided eye along the pathway between disease and distal genotype—for genetic risk factors of ASD. Using a genome-by-metabolome-wide association framework, we were able to evaluate potential associations between a broad spectrum of small molecules, encompassing both endogenous metabolites and exogenous compounds, and common genetic variants across the whole genome. Our primary objective was to detect and prioritize genotype-metabolite associations unique to the AGRE cohort, not present in the PLHD cohort. We identified a significant association between plasma glutamine levels and specific genetic variants within the *NLGN1* gene, which was further bolstered by the correlation between plasma glutamine levels and the severity of repetitive and restricted behaviours (RRBs). These findings suggest that plasma glutamine levels could reflect *NLGN1*'s genetic influence on the severity of RRBs in individuals with ASD, potentially serving as an endophenotype.

The *NLGN* gene family, including *NLGN1*, are crucial post-synaptic cell-adhesion molecules, establish physical interactions with pre-synaptic NRXNs and post-synaptic scaffolding proteins such as SHANK proteins. These NLGN-NRXN interactions are essential for synaptogenesis and synaptic maintenance, thereby

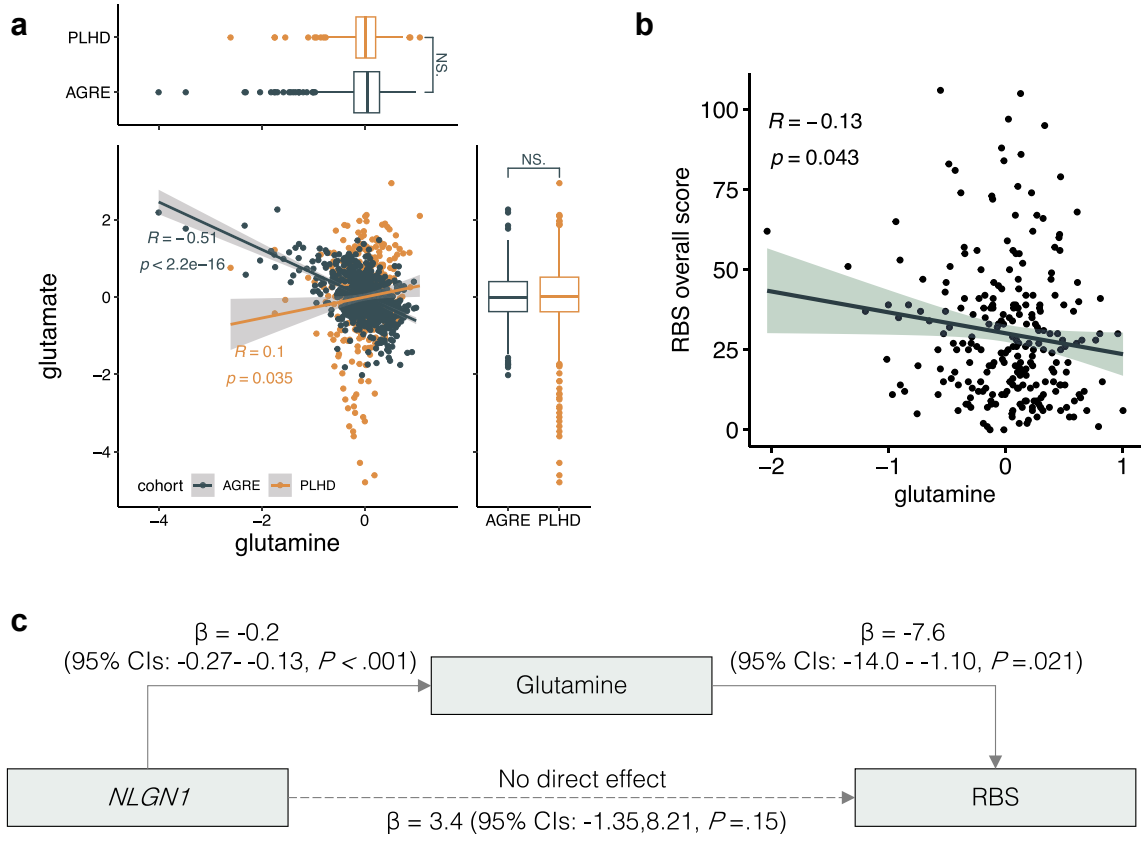

Fig. 5: **The correlation between glutamine and glutamate levels, and association between *NLGN1* and Repetitive Behaviour Scale (RBS) total score.** (a) The correlation between glutamine and glutamate levels in the AGRE and PLHD cohort. The boxplots in the top and right panels depict the distribution of normalized glutamine and glutamate levels in both cohorts. (b) The correlation between glutamine levels and RBS total score. (c) A diagram illustrating the impact of *NLGN1* and plasma glutamine levels on the RBS total scores.

significantly contribute to the proper formation and operation of neural circuits in the human brain.[54,55] Mutations in these genes can dramatically disrupt neural circuit formation and functioning, potentially leading to neurodevelopmental disorders.[56] Members of the *NLGN* gene family have distinct patterns of expression and roles. For instance, NLGN1 is exclusively localized to the postsynaptic dendritic membrane of excitatory neurons, whereas NLGN2 is typically expressed at GABAergic inhibitory synapses. NLGN3 has a broader distribution, found at both glutamatergic and GABAergic synapses, and NLGN4X is detected at excitatory synapses.[57–59] Genetic variants in *NLGN1*, including copy number variations (CNVs), rare inherited, and *de novo* variants, have been identified in individuals with ASD.[60–62] *De novo* variants in the *NLGN3* and *NLGN4* genes have also been discovered in individuals with ASD.[63–65]

Notably, the *NLGN* gene family, through its modulation of the cortical E-I balance, may significantly contribute to the pathogenesis of ASD.[56,66] Animal models, such as *Nlgn1* knock-out (Nlgn1-KO) mice, have provided valuable insights into the effects of these genetic alterations. These mice exhibited deficits in spatial learning and memory, linked to disrupted N-methyl-D-aspartic acid (NMDA) receptor signalling in excitatory neurons. Moreover, these mice showed a marked increase in repetitive grooming behaviours, a potential ASD-related RRB phenotype. This behaviour was correlated with a reduced NMDA/AMPA ratio at corticostriatal synapses, and intriguingly, the excessive grooming phenotype could be alleviated with the administration of an NMDA receptor partial coagonist.[67] Knock-in mice carrying the human missense variant p. Pro89Leu in the *NLGN1* gene exhibited social deficits and impaired spatial memory—traits often associated with ASD.[68] In particular, mutations in *NLGN4* are most frequently found and exhibit high penetrance.[59,69] Collectively, these observations underscore the potential contribution of *NLGN* genes to the genetic liability of ASD, and the role of *Nlgn1* in modulating excitatory synaptic transmission, potentially underlying ASD pathophysiology and offering a therapeutic target for RRBs.

By utilizing an untargeted methodology, we identified a significant association between *NLGN1* and plasma glutamine levels within the AGRE cohort, an association not observed in the PLHD cohort. A significant inverse correlation was also observed between plasma glutamine and glutamate levels within the AGRE cohort, a correlation not found in the PLHD cohort. Though speculative, these findings suggest possible aberrations in the glutamate–glutamine cycle in a subset of individuals with ASD, which might be linked to genetic variants in *NLGN1*. Furthermore, the correlation observed between plasma glutamine levels—rather than glutamate levels—and the severity of repetitive and restricted behaviours (RRBs) highlights the significance of our findings. Nonetheless, while our findings provide valuable insights into the potential role of *NLGN1* and its association with RRBs, further research is needed to validate these findings.

In the CNS, the glutamate-glutamine cycle is essential for the metabolism and recycling of glutamate between neurons and astrocytes. Disruptions in this cycle, potentially due to genetic or environmental risk factors, could lead to an imbalance between glutamate and GABA levels. These changes may lead to an imbalance between excitatory and inhibitory neuronal activity, potentially contributing to the underlying pathophysiology of ASD. This E-I imbalance, often linked to alterations in synaptic function or connectivity that enhance excitation or reduce inhibition,[70] is proposed to contribute to core ASD symptoms, including impaired social interaction, communication deficits, and repetitive behaviours.[71] Our findings suggest a potential role for *NLGN1* in modulating the glutamine-glutamate cycle, and thereby, the E-I balance. This aligns with theories suggesting an E-I imbalance underlies the ASD pathophysiology.[38,72] Further research is warranted to investigate the direct involvement of NLGN1 in the glutamine–glutamate cycle and its potential influence on the RRB phenotype.

Glutamine is the most abundant amino acid in plasma and cerebrospinal fluid (CSF), and plasma glutamine concentration affects brain glutamine concentration.[73] The concentration of glutamine in the posterior cingulate cortex, as measured by magnetic resonance spectrometry, was correlated with the concentration of glutamine in plasma, but not for glutamate.[74] Glutamine is a precursor to glutamate and GABA, which are the primary excitatory and inhibitory neurotransmitters in the mammalian brain. In addition, outside the brain, glutamine also plays a crucial role in supporting intestinal cell energy, regulating the immune system function, and maintaining acid-base balance in the body.[75] Glutamine and glutamate levels have been studied in the brain and blood for various psychiatric disorders, including schizophrenia and ASD.[76] Reduced glutamate/glutamine ratios were reported using *in vivo* proton magnetic resonance spectroscopy for ASD,[34] and plasma glutamine levels were significantly reduced in ASD compared to controls.[77] A post-mortem study measuring glutamine-glutamate cycle processing enzymes showed a significantly lower level of kidney-type glutaminase in the anterior cingulate cortex of post-mortem brain samples from individuals with ASD compared to those from controls.[78]

Glutamatergic dysfunction, leading to an imbalance of cortical cellular excitation to inhibition, is posited as a common mechanism in ASD pathophysiology.[76] Magnetic resonance imaging studies identified a normal GABA but decreased glutamate concentration in the striatum of individuals with ASD, a pattern correlating with social symptom severity, pointing towards subcortical dysfunctions in ASD.[35] Post-mortem examinations of ASD brains have unveiled changes in the density and distribution of GABAergic interneurons, indicating potential abnormalities in inhibitory interneurons.[79] Similarly, a disruption in the correlation between GABA levels and perceptual dynamics has been reported in individuals with ASD, suggesting elevated excitatory-to-inhibitory signalling in the visual cortex.[80] In a pioneering study, Deisseroth and colleagues employed bistable optogenetic modulation to demonstrate that E-I imbalance in the mouse prefrontal cortex could induce social dysfunction without motor abnormalities, and that restoring E-I balance through the elevation of inhibitory signals could ameliorate social deficits.[71] These findings suggest that the social dysfunction phenotype in ASD may be associated with the E-I imbalance in the medial prefrontal cortex (mPFC), attributed to an excess of excitatory neurons. Altered E-I balance may also be linked with potential gender-specific differences in ASD. A study using the Hurst exponent (H), a time-series metric derived from fMRI BOLD signals as an indicator of synaptic E-I ratio changes, reported decrease in H (suggesting increased excitation) in the mPFC of males with ASD but not females.[81]

Our study has several limitations. Firstly, plasma sample collection, preparation, and biobanking might have contributed to global differences in metabolomics profiling. The AGRE and PLHD samples were collected during different time periods, leading to variations in the duration of storage in freezers between the two collections. As a result, we could not perform a case-control comparison of feature levels. Secondly, the cohorts were genotyped using different methods—WGS for AGRE and a genotyping microarray for PLHD. Therefore, we limited our association analysis to common variants that were confidently profiled by both platforms. Thirdly, the lack of specific measurements of autistic traits, such as social impairment and repetitive and restricted behaviours (RRB), in the PLHD cohort presented another limitation. Even though none of individuals in PLHD were diagnosed with ASD, variations or subclinical traits related to ASD could still exist.

Furthermore, none of the participants in the PLHD cohort underwent neuropsychological evaluations, which limited our ability to fully compare behavioural and genetic correlations across the ASD and control cohorts. This could introduce potential confounders, as unassessed traits or neuropsychological factors could influence the observed associations unique to the AGRE cohort.

Despite these limitations, our study presents several notable strengths and innovative aspects that significantly advance our understanding of the potential mechanisms underpinning the genetic risk factors associated with ASD. Firstly, our research is distinguished by our application of a genome-by-metabolome-wide association study (GxMWAS). This method, incorporating high-resolution metabolomics, parallels the transformative genome-wide association analysis (GWAS) framework, enhancing our capacity to discover candidate genes and endophenotypes in an unbiased manner. Secondly, we utilized metabolites as endophenotypes to better understand the underlying mechanisms of ASD genetic risk factors. Thirdly, we successfully discovered the association between *NLGN1* and glutamine, and correlated plasma glutamine levels with one of the core symptoms of ASD. Fourthly, our inclusion of a non-ASD cohort strengthens the specificity of the observed association between *NLGN1* and glutamine in relation to ASD, emphasizing the potential clinical significance of plasma glutamine levels in ASD. Finally, we employed an unprecedented sample size in metabolome research, which enhances the statistical power of our findings and uncovers relationships that may have been missed in smaller-scale studies. As a result, our study sets a new standard for future metabolome-wide investigations in the field. These unique aspects of our study have facilitated the identification of previously unknown relationships between genetic variants, plasma metabolite levels, and ASD severity, providing valuable insights into potential therapeutic targets and personalized treatment strategies for individuals with ASD. We believe that our study significantly contributes to the field, paving the way for future research and enhancing our understanding of ASD and other complex disorders.

In conclusion, our study has unveiled a unique association between *NLGN1* haplotypes and plasma glutamine levels in individuals with ASD and their family members. We have also detected a significant correlation between plasma glutamine levels and the severity of RRBs, a core symptom domain of ASD. While we did not find a direct association between *NLGN1* variants and RRB scores, our findings suggest that plasma glutamine levels could potentially serve as an endophenotype, mediating the genetic influence of *NLGN1* on the severity of RRBs. This study highlights the potential role of E-I imbalance in the pathophysiology of ASD and proposes plasma glutamine levels as a potential biomarker for identifying specific ASD subgroups. These identified subgroups could benefit from interventions targeting the glutamatergic signalling pathway in clinical trials.[38] Future research should aim to confirm the role of *NLGN1* in the glutamine-glutamate cycle and its impact on RRBs. Additionally, further investigations into alterations in the E-I balance could provide invaluable insights for the development of novel therapeutic interventions for ASD.

Contributors

Study concept and design: SWK, DIW, DPJ; Data acquisition, analysis, or interpretation: MRS, DIW, I-HL, SWK; Verification of genotype data: I-HL, YL, SWK; Verification of metabolomics data: DIW, MRS, YL, SWK; Verification of phenotype data: YL, SWK; Validation of integration of genotype, metabolomic and phenotype data: I-HL, SWK; Drafting of the manuscript: I-HL, MRS, DIW, SWK; Critical revision of the manuscript: I-HL, MRS, DIW, KDM, DPJ, SWK; Statistical analysis: I-HL, DIW, SWK; Funding acquisition: SWK, KDM; Administrative, technical, or material support: YL; Supervision: I-HL, MRS, YL. All authors were involved in the development of the primary manuscript and have read and approved the final version.

Data sharing statement

The plasma metabolome profiling data generated in this study are available at Metabolomic Workbench (https://www.metabolomicsworkbench.org/) under project IDs PR001496 (for AGRE cohort) and PR001495 (for the PLHD cohort). WGS data of the AGRE cohort can be obtained directly from MSSNG and iHART. Detailed phenotype information for AGRE participants can be obtained directly from Autism Speaks. Due to the sensitive nature of the data, extracts from BCH EHR database and genotype data for the PLHD participants are not available. Summary statistics for the 166 identified metabolites are available in the GWAS catalogue. Summary statistics for the other features are available upon request.

Declaration of interests

All authors have no potential conflicts of interests to disclose.

Acknowledgements

This work was supported by the National Institute of Mental Health (grant R01MH107205 to Dr. Kong). This work was additionally supported in part by the National Institute of Health (grant numbers: U01TR002623, R24OD024622, OT2OD032720, and R01NS129188), and by the PrecisionLink Biobank for Health Discovery at Boston Children's Hospital.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104746.

References

1 Maenner MJ, Shaw KA, Baio J, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveill Summ*. 2020;69(4):1–12.
2 Christensen DL, Baio J, Van Naarden Braun K, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years–autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveill Summ*. 2016;65(3):1–23.
3 Tick B, Bolton P, Happe F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry*. 2016;57(5):585–595.
4 Antaki D, Guevara J, Maihofer AX, et al. A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nat Genet*. 2022;54(9):1284–1292.

5 Abrahams BS, Arking DE, Campbell DB, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*. 2013;4(1):36.
6 Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol*. 2014;10(2):74–81.
7 De Rubeis S, He X, Goldberg AP, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515(7526):209–215.
8 Chen J, Yu S, Fu Y, Li X. Synaptic proteins and receptors defects in autism spectrum disorders. *Front Cell Neurosci*. 2014;8:276.
9 Fu JM, Satterstrom FK, Peng M, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat Genet*. 2022;54(9):1320–1331.
10 Myers SM, Challman TD, Bernier R, et al. Insufficient evidence for "Autism-Specific" genes. *Am J Hum Genet*. 2020;106(5):587–595.
11 Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856–860.
12 King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet*. 2019;15(12):e1008489.
13 Spratt PWE, Ben-Shalom R, Keeshen CM, et al. The autism-associated gene Scn2a contributes to dendritic excitability and synaptic function in the prefrontal cortex. *Neuron*. 2019;103(4):673–685.e5.
14 American Psychiatric A. *Diagnostic and statistical manual of mental disorders*. 5 ed. Washington, DC: American Psychiatric Association; 2013.
15 Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder. *Lancet*. 2018;392(10146):508–520.
16 McClellan J, King MC. Genetic heterogeneity in human disease. *Cell*. 2010;141(2):210–217.
17 Gottesman II, Shields J. Genetic theorizing and schizophrenia. *Br J Psychiatry*. 1973;122(566):15–30.
18 Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003;160(4):636–645.
19 Anderson GM, Gutknecht L, Cohen DJ, et al. Serotonin transporter promoter variants in autism: functional effects and relationship to platelet hyperserotonemia. *Mol Psychiatry*. 2002;7(8):831–836.
20 Modahl C, Green L, Fein D, et al. Plasma oxytocin levels in autistic children. *Biol Psychiatry*. 1998;43(4):270–277.
21 Belmonte MK, Gomot M, Baron-Cohen S. Visual attention in autism families: 'unaffected' sibs share atypical frontal activation. *J Child Psychol Psychiatry*. 2010;51(3):259–276.
22 Spezio ML, Adolphs R, Hurley RS, Piven J. Abnormal use of facial information in high-functioning autism. *J Autism Dev Disord*. 2007;37(5):929–939.
23 Alarcon M, Abrahams BS, Stone JL, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet*. 2008;82(1):150–159.
24 Strauss KA, Puffenberger EG, Huentelman MJ, et al. Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med*. 2006;354(13):1370–1377.
25 Lee IH, Koelliker E, Kong SW. Quantitative trait locus analysis for endophenotypes reveals genetic substrates of core symptom domains and neurocognitive function in autism spectrum disorder. *Transl Psychiatry*. 2022;12(1):407.
26 Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol*. 2016;17(7):451–459.
27 Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543–550.
28 Rinschen MM, Ivanisevic J, Giera M, Siuzdak G. Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol*. 2019;20(6):353–367.
29 Lotta LA, Pietzner M, Stewart ID, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. *Nat Genet*. 2021;53(1):54–64.
30 Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. Computational metabolomics: a framework for the million metabolome. *Chem Res Toxicol*. 2016;29(12):1956–1975.
31 Jones DP. Sequencing the exposome: a call to action. *Toxicol Rep*. 2016;3:29–45.
32 Chung MK, Smith MR, Lin Y, et al. Plasma metabolomics of autism spectrum disorder and influence of shared components in proband families. *Exposome*. 2021;1(1):osab004.
33 Lee IH, Smith MR, Yazdani A, et al. Comprehensive characterization of putative genetic influences on plasma metabolome in a pediatric cohort. *Hum Genomics*. 2022;16(1):67.
34 Horder J, Lavender T, Mendez MA, et al. Reduced subcortical glutamate/glutamine in adults with autism spectrum disorders: a [(1)H]MRS study. *Transl Psychiatry*. 2013;3(7):e279.
35 Horder J, Petrinovic MM, Mendez MA, et al. Glutamate and GABA in autism spectrum disorder-a translational magnetic resonance spectroscopy study in man and rodent models. *Transl Psychiatry*. 2018;8(1):106.
36 Etherton MR, Blaiss CA, Powell CM, Sudhof TC. Mouse neurexin-1alpha deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. *Proc Natl Acad Sci U S A*. 2009;106(42):17998–18003.
37 Verpelli C, Dvoretskova E, Vicidomini C, et al. Importance of Shank3 protein in regulating metabotropic glutamate receptor 5 (mGluR5) expression and signaling at synapses. *J Biol Chem*. 2011;286(40):34839–34850.
38 Sohal VS, Rubenstein JLR. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol Psychiatry*. 2019;24(9):1248–1257.
39 Lee E, Lee J, Kim E. Excitation/inhibition imbalance in animal models of autism spectrum disorders. *Biol Psychiatry*. 2017;81(10):838–847.
40 Geschwind DH, Sowinski J, Lord C, et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet*. 2001;69(2):463–466.
41 Bastarache L. Using phecodes for research with the electronic health record: from PheWAS to PheRS. *Annu Rev Biomed Data Sci*. 2021;4:1–19.
42 Lingren T, Chen P, Bochenek J, et al. Electronic health record based algorithm to identify patients with autism spectrum disorder. *PLoS One*. 2016;11(7):e0159621.
43 Ritz B, Yan Q, Uppal K, et al. Untargeted metabolomics screen of mid-pregnancy maternal serum and autism in offspring. *Autism Res*. 2020;13(8):1258–1269.
44 Liu KH, Nellis M, Uppal K, et al. Reference standardization for quantification and harmonization of large-scale metabolomics. *Anal Chem*. 2020;92(13):8836–8844.
45 Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol*. 2014;48(4):2097–2098.
46 Trost B, Thiruvahindrapuram B, Chan AJS, et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell*. 2022;185(23):4409–4427.e18.
47 Ruzzo EK, Perez-Cano L, Jung JY, et al. Inherited and de novo genetic risk for autism impacts shared networks. *Cell*. 2019;178(4):850–866.e26.
48 Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290–299.
49 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
50 Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res*. 2018;27(2):e1608.
51 Yang J, Ferreira T, Morris AP, Medland SE. Genetic Investigation of ATC, Replication DIG, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369–375. S1-S375.
52 Hayes AF, Rockwood NJ. Regression-based statistical mediation and moderation analysis in clinical research: observations, recommendations, and implementation. *Behav Res Ther*. 2017;98:39–57.
53 Little J, Higgins JP, Ioannidis JP, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement. *Hum Genet*. 2009;125(2):131–151.
54 Dean C, Dresbach T. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci*. 2006;29(1):21–29.
55 Varoqueaux F, Aramuni G, Rawson RL, et al. Neuroligins determine synapse maturation and function. *Neuron*. 2006;51(6):741–754.

56 Sudhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature*. 2008;455(7215):903–911.
57 Chubykin AA, Atasoy D, Etherton MR, et al. Activity-dependent validation of excitatory versus inhibitory synapses by neuroligin-1 versus neuroligin-2. *Neuron*. 2007;54(6):919–931.
58 Bemben MA, Shipman SL, Nicoll RA, Roche KW. The cellular and molecular landscape of neuroligins. *Trends Neurosci*. 2015;38(8):496–505.
59 Marro SG, Chanda S, Yang N, et al. Neuroligin-4 regulates excitatory synaptic transmission in human neurons. *Neuron*. 2019;103(4):617–626.e6.
60 Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459(7246):569–573.
61 Tejada MI, Elcoroaristizabal X, Ibarluzea N, Botella MP, de la Hoz AB, Ocio I. A novel nonsense homozygous variant in the NLGN1 gene found in a pair of monozygotic twin brothers with intellectual disability and autism. *Clin Genet*. 2019;95(2):339–340.
62 O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012;485(7397):246–250.
63 Jamain S, Quach H, Betancur C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*. 2003;34(1):27–29.
64 Laumonnier F, Bonnet-Brilhault F, Gomot M, et al. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet*. 2004;74(3):552–557.
65 Yan J, Oliveira G, Coutinho A, et al. Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol Psychiatry*. 2005;10(4):329–332.
66 Chih B, Engelman H, Scheiffele P. Control of excitatory and inhibitory synapse formation by neuroligins. *Science*. 2005;307(5713):1324–1328.
67 Blundell J, Blaiss CA, Etherton MR, et al. Neuroligin-1 deletion results in impaired spatial memory and increased repetitive behavior. *J Neurosci*. 2010;30(6):2115–2129.
68 Nakanishi M, Nomura J, Ji X, et al. Functional significance of rare neuroligin 1 variants found in autism. *PLoS Genet*. 2017;13(8):e1006940.
69 Nguyen TA, Lehr AW, Roche KW. Neuroligins and neurodevelopmental disorders: X-linked genetics. *Front Synaptic Neurosci*. 2020;12:33.
70 Nelson SB, Valakh V. Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron*. 2015;87(4):684–698.
71 Yizhar O, Fenno LE, Prigge M, et al. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*. 2011;477(7363):171–178.
72 Rubenstein JL, Merzenich MM. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav*. 2003;2(5):255–267.
73 Bagga P, Behar KL, Mason GF, De Feyter HM, Rothman DL, Patel AB. Characterization of cerebral glutamine uptake from blood in the mouse brain: implications for metabolic modeling of 13C NMR data. *J Cereb Blood Flow Metab*. 2014;34(10):1666–1672.
74 Takado Y, Sato N, Kanbe Y, et al. Association between brain and plasma glutamine levels in healthy young subjects investigated by MRS and LC/MS. *Nutrients*. 2019;11(7):1614.
75 Altman BJ, Stine ZE, Dang CV. From Krebs to clinic: glutamine metabolism to cancer therapy. *Nat Rev Cancer*. 2016;16(10):619–634.
76 Nisar S, Bhat AA, Masoodi T, et al. Genetics of glutamate and its receptors in autism spectrum disorder. *Mol Psychiatry*. 2022;27(5):2380–2392.
77 Shimmura C, Suda S, Tsuchiya KJ, et al. Alteration of plasma glutamate and glutamine levels in children with high-functioning autism. *PLoS One*. 2011;6(10):e25340.
78 Shimmura C, Suzuki K, Iwata Y, et al. Enzymes in the glutamate-glutamine cycle in the anterior cingulate cortex in postmortem brain of subjects with autism. *Mol Autism*. 2013;4(1):6.
79 Hashemi E, Ariza J, Rogers H, Noctor SC, Martinez-Cerdeno V. The number of parvalbumin-expressing interneurons is decreased in the prefrontal cortex in autism. *Cereb Cortex*. 2017;27(3):1931–1943.
80 Robertson CE, Ratai EM, Kanwisher N. Reduced GABAergic action in the autistic brain. *Curr Biol*. 2016;26(1):80–85.
81 Trakoshis S, Martinez-Canada P, Rocchi F, et al. Intrinsic excitation-inhibition imbalance affects medial prefrontal cortex differently in autistic men versus women. *Elife*. 2020;9:e55684.