



OPEN Shared rare genetic variants in multiplex autism families suggest a social memory gene under selection

Kang Seon Lee^{1,7}, Taeyeop Lee^{1,2,3,4,7}, Mujun Kim^{5,7}, Elizaveta Ignatova¹, Hyo-Jeong Ban¹, Min Kyung Sung¹, Younghoon Kim¹, Youn-Jae Kim⁶, Jin-Hee Han⁵✉ & Jung Kyoony Choi¹✉

Autism spectrum disorder (ASD) affects up to 1 in 59 children, and is one of the most common neurodevelopmental disorders. Recent genomic studies have highlighted the role of rare variants in ASD. This study aimed to identify genes affected by rare variants shared by siblings with ASD and validate the function of a candidate gene *FRRS1L*. By integrating the whole genome sequencing data of 866 multiplex families from the Hartwell Foundation's Autism Research and Technology Initiative and Autism Speaks MSSNG project, we identified rare variants shared by two or more siblings with ASD. Using shared rare variants (SRVs), we selected candidate genes for ASD. Gene prioritization by evolutionary features and expression alterations on autism identified *FRRS1L* in two families, including one with impaired social behaviors. One variant in this family was 6 bp away from human-specific trinucleotide fixation. Additionally, CRISPR/Cas9 experiments demonstrated downregulation by a family variant and upregulation by a fixed site. Population genetics further demonstrated that upregulation of this gene has been favored during human evolution. Various mouse behavioral tests showed that *Frrs1l* knockout specifically impairs social novelty recognition without altering other behavioral phenotypes. Furthermore, we constructed humanized mice by introducing human sequences into a mouse genome. These knockin mice showed improved abilities to retain social memory over time. The results of our population genetic and evolutionary analyses, behavior experiments, and genome editing propose a molecular mechanism that may confer a selective advantage through social memory enhancement and may cause autism-related social impairment when disrupted in humans. These findings highlight the role of *FRRS1L*, the AMPA receptor subunit, in social behavior and evolution.

Keywords Autism spectrum disorder, Shared rare variant, Social novelty, Social memory, *FRRS1L*, AMPA receptor

Autism spectrum disorder (ASD) is characterized by persistent abnormalities in social interactions and communication and restricted, repetitive behavior, interests, or activities¹. According to a recent study of the National Health Interview Survey, ASD affects up to 1 in 59 children in the United States². Currently, the prevalence of ASDs is steadily increasing.

The genetic architecture of autism is complex and highly heterogeneous, and the allelic spectrum influencing the variability of traits is a key component of it³. Over the past decade, the genetic architecture of ASD has become clearer, and numerous risk variants and ASD-related genes have been identified using genome-wide microarrays⁴, whole-exome sequencing^{5,6} and whole-genome sequencing^{7,8}. The majority of the discovered genes are based on de novo or rare variants found in the coding regions of simplex families^{9,10}. In addition to

¹Department of Bio and Brain Engineering, KAIST, Daejeon 34141, Republic of Korea. ²Graduate School of Medical Science and Engineering, KAIST, Daejeon 34141, Republic of Korea. ³Department of Psychiatry, University of Ulsan College of Medicine, Asan Medical Center, Seoul 05505, Republic of Korea. ⁴Translational Biomedical Research Group, Asan Institute for Life Science, Asan Medical Center, Seoul 05505, Republic of Korea. ⁵Department of Biological Sciences, KAIST, Daejeon 34141, Republic of Korea. ⁶Specific Organs Cancer Branch, National Cancer Center, Gyeonggi 10408, Republic of Korea. ⁷These authors contributed equally: Kang Seon Lee, Taeyeop Lee and Mujun Kim. ✉email: han.jinhee@kaist.ac.kr; jungkyoon@kaist.ac.kr

the genes from the simplex families, rare variants that are shared in multiplex families may confer a significant risk of ASD¹¹.

Rare variants in the noncoding genome contribute to ASD risk through alterations in gene regulatory mechanisms^{12,13}. Variants in the regulatory noncoding region can activate or inactivate gene transcription by gaining or losing transcription factor (TF) binding sites¹⁴. In particular, when a conserved genomic region with high divergence in humans harbors a rare variant, human-specific cognition and social behavior may be disrupted¹⁵. These regions, which are known as human-accelerated regions, show enrichment for rare variants, especially in individuals with ASD^{15–17}. Findings from the human-accelerated regions highlight the importance of considering evolutionary signals when investigating the regulatory noncoding genome.

In this study, we present shared rare variants (SRVs) derived from multiple whole-genome sequences, which enabled us to discover candidate genes. Consecutive mouse experiments supported our findings and informed the function of the target gene, which helped uncover the biological mechanisms affecting the clinical consequences of gene function disruption.

Methods

ASD family data acquisition and processing

As part of a collaboration established between the Autism Speaks the MSSNG Project and Hartwell Foundation's Autism Research and Technology Initiative (iHART), whole-genome sequencing and detailed phenotypic information of individuals affected with ASD and their families are available to the research community^{8,11,18,19}. Clinical data, including race/ethnicity and the Vineland Adaptive Behavior Scales, were obtained from the MSSNG and iHART online portal (<https://research.mss.ng/> and <http://www.ihart.org>, respectively). Access to the MSSNG and iHART data was obtained by completing data access agreements, and the approval by the relevant institutions are required before obtaining the data. Since we downloaded anonymous data from the MSSNG and iHART consortium, it was impossible to identify individuals and obtain additional consent from all subjects and/or their legal guardians. This study, including the data acquisition process, was approved by the Institutional Review Board of KAIST (IRB-16-93 and IRB-19-439), and informed consent requirement was waived.

Using the application programming interfaces from Cloud Life Sciences (formerly Google Genomics) in Google Cloud and Amazon Web Services, we downloaded the universal variant call format (VCF) files of the MSSNG and iHART data for multiplex families, in which parents were unaffected and all children were affected. In this study, we analyzed the VCF files of ASD children from multiplex families in the MSSNG and iHART databases. We used 964 samples from 445 multiplex families from the MSSNG database and 991 samples from 421 multiplex families from the iHART database. Overlapping families were removed. The iHART VCF files were lifted over to hg19/GRCh37 from hg38/GRCh38. We identified homozygous or heterozygous rare variants that were shared by all affected children in a multiplex family. To reduce the error rate of the variants, we additionally filtered variants by Filtered Depth < 20 and Genotype Quality < 80. Rare variants were defined using the Genome Aggregation Database (gnomAD), which included 15,708 whole-genome sequences (<https://gnomad.broadinstitute.org/>), and the 1000 Genomes Project, which included 2504 whole-genome sequences²⁰. We selected variants with a minor allele frequency of < 0.01 across the gnomAD and 1000 genomes. Highly polymorphic regions were excluded, enabling alternate sequence representations. GRCh37 contains nine alternate sequences in chr6: 28,477,796–33,448,353, chr17: 43,384,863–44,913,631, and chr4: 69,170,076–69,878,206. We confirm that all methods involved in this study were carried out in accordance with the ethical guidelines established by the national and institutional committees overseeing human experimentation, as well as the Helsinki Declaration.

Positive selection test and scoring

We searched SRV-flanking sequences for evolutionary sequence divergence sites. When derived and ancestral alleles coexisted in a specific sequence, that site was defined as an evolutionary sequence divergence site. When defining divergence sites, we considered cases in which a human sequence contained the derived allele, and at least two primates carried the ancestral allele. Derived and ancestral alleles were assessed in primates, including humans, chimpanzees, gorillas, orangutans, and macaques, and their sequences were obtained from the UCSC genome browser (<https://genome.ucsc.edu>).

For gene prioritization, we computed gene-level scores for the evolutionary sequence divergence sites. We counted the number of evolutionary sequence divergence sites located at 10 bp upstream and 10 bp downstream of each SRV. We then obtained an average count for all the SRVs in each gene. This average count served as an 'Evolutionary sequence divergence score' in Fig. 2B.

We also searched the SRV-flanking sequences for human-specific fixation events. Human-specific fixation events were defined as variants whose derived allele frequency was > 99% among the 2504 samples from the 1000 Genomes Project. Specifically, an allele was regarded as derived when only humans carried a sequence different from the other primates (chimpanzees, gorillas, orangutans, and macaques). We tested whether the fixed sites were accompanied by quantitative signs of positive selection. We used the 1000 Genomes samples for these analyses. When positive selection increases the frequency of a favored allele, neighboring neutral sequences are swept through the population along with the selected variant. This process called selective sweep causes an excess of linkage disequilibrium (LD), a decrease in the level of genetic diversity, and skew of the site frequency spectrum²¹. As the LD breaks down quickly over time, tests based on extended haplotype provide weak power to detect ancient sweeps, leading to fixation, and are useful only for detecting recent, partial selective sweeps²². Therefore, we tested for a reduction in nucleotide diversity and bias in the allele frequency spectrum.

Nucleotide diversity per site was estimated as follows: $\pi = 2p(1-p)n/(n-1)$, where p is the frequency of the non-reference allele, and n is the number of chromosomes in the examined population. We computed

the average π per site by summing π for all nucleotides in a 1-kb running window and dividing it by the total number of bases within the window because $\pi = 0$ at the invariant sites.

To measure the skew of the allele frequency spectrum, we performed the Tajima's D test²³. In the 1-kb running windows, the number of segregating sites and nucleotide differences between pairs sampled (pairwise differences) were obtained. Under neutrality, these two measures have equal expectations, and Tajima's D will be close to 0. Negative values of the Tajima's D indicate an excess of rare variations, which is consistent with positive selection. The statistic was calculated with the following parameters: n , number of chromosomes; S_n , number of polymorphic sites observed; and p_i , major allele frequency of the i th variant. With these parameters, the Tajima's D was obtained as follows:

$$D = \frac{\pi - \theta_s}{\sqrt{\text{Var}(\pi - \theta_s)}},$$

where

$$\pi = \frac{n}{n-1} \sum_{i=1}^{S_n} 2p_i(1-p_i)$$

and

$$\theta_s = \frac{S_n}{\sum_{i=1}^{n-1} \frac{1}{i}}.$$

To detect recent sweeps, we performed the integrated Haplotype Score (iHS) test²⁴, as an LD-based method. The iHS test was developed based on the extended haplotype homozygosity (EHH) statistic²², under the principle that, under positive selection, an allele may sweep to high frequency rapidly enough that long-distance association with neighboring polymorphisms will not have sufficient time to decay by recombination²⁵. The EHH measures the decay of identity of haplotypes that carry a specified core allele. Extreme iHS scores ($|iHS| > 2$) indicate that haplotypes with specified allele are longer than those with background alleles. We used a bioinformatics workflow for detecting signatures²⁶ to compute the iHS and determine the selected allele for individual variants. For gene prioritization, we computed the proportion of genomic loci with extreme iHS scores mapping to each gene. This measure served as the 'haplotype-based positive selection score' in Fig. 2B.

Other data analyses

To compare gene expression in human tissues for the candidate genes, we obtained the Genotype-Tissue Expression (GTEx) RNA-seq data for 54 types of human tissue expression from the GTEx portal (GTEx Analysis V8 release)²⁷. Using about 10,000 gene expression levels, we calculated the average value of the gene expression level for each of the 54 tissues, and z-score normalization was then performed for each gene to compare the tissue-specific expression patterns of each gene.

To assess the genomic context of the SRVs, we used ChIPSeeker. The genomic contexts were categorized as promoter (± 3 kb of any transcription start sites), 5'-untranslated region (UTR), 3'-UTR, exon, intron, and distal intergenic²⁸.

For the expression quantitative loci (eQTL) analyses, we used the GTEx²⁷ data portal (<https://www.gtexportal.org>). Information about rs2439649 from a genome-wide association study (GWAS) for educational attainment²⁹ was obtained from the GWAS catalog (<https://www.ebi.ac.uk/gwas>).

Transcriptome data in ASD versus normal brains were acquired from a previously published dataset³⁰. We used cases with chromosome 15q11.2-13.1 duplication syndrome (dup15q), one of the most common and penetrant forms of ASD. We compared the expression level of the *FRRS1L* gene in the developing brains of humans and rhesus macaques based on processed mRNA sequencing data from the cortex regions of prenatal and postnatal brains³¹. The *FRRS1L* expression pattern across various cell types in human prenatal brains was obtained from the PsychENCODE consortium³². A gene coexpression network was constructed using the Bioconductor package 'weighted gene correlation network analysis'³³ based on human fetal brain transcriptome data obtained from the BrainSpan Atlas³⁴.

CRISPR/Cas9 genome editing

To generate a specific sequence change, we used the two-step sequence editing strategy 'CORRECT'³⁵ based on homology-directed repair (HDR). In the first step, we constructed two wildtype single guide RNA (WT-sgRNA) molecules, each flanking the two target sites (C-to-T substitution at chr9: 111,924,882 and CC-to-TT substitution at chr9: 111,924,875–111,924,876) using RNA-guided engineered nucleases. Possibility of off-target effects were predicted using in silico tools (<http://www.rgenome.net>)^{36,37}. These WT-sgRNAs were cloned into pSpCas9(BB)-2A-GFP (PX458, Addgene, #48138). The pSpCas9(BB)-2A-GFP vector with the WT-sgRNA and HDR templates was transfected into SH-SY5Y cells using the Neon[®] Transfection System kit (Thermo Fisher Scientific). Additionally, an HDR template containing the intended mutation (M) and CRISPR/Cas9-blocking mutations (B) (MB HDR template) was introduced. After 48 h, transfected GFP-positive cells were isolated using fluorescence-activated cell sorting (FACS). The isolated cells harboring the MB HDR template were cultured for two weeks for re-transfection. In the second step, we constructed re-single guide RNA (re-sgRNA) for the respective MB HDR target site. The cells from the first step were then transfected with the pSpCas9(BB)-2A-GFP vector with the re-sgRNA and HDR template containing only the intended mutation (M). After 48 h, the transfected GFP-positive cells were individually isolated using FACS. The single-cell clones were individually

cultured in single wells for one month. Genomic DNA was extracted from the cultured cells using the DNeasy Blood and Tissue kit (Qiagen) and amplified by polymerase chain reaction (PCR) using Hifi Hot Start (KAPA). The obtained PCR products were sequenced in forward and reverse orientations using Sanger sequencing. The RNA was extracted using the RNeasy Plus mini kit (QIAGEN), and complementary DNA (cDNA) was synthesized from total RNA using SuperScript IV VILO Master Mix (Invitrogen). Quantitative real-time reverse-transcription (qRT) PCR was performed using the SYBR Green PCR Master Mix (Applied Biosystems) on the QuantStudio[®] 5 RT-PCR System (Applied Biosystems). The *FRRS1L* expression levels were measured using qRT-PCR and normalized by the *GAPDH* levels. Each of the three independent experiments was performed with three technical replicates. Sequences of the PCR primers and CRISPR sgRNA used in this work are provided in Supplementary Tables 1–4.

We used the SH-SY5Y cell line obtained from the American Type Culture Collection. The cells were cultured in complete Dulbecco's modified Eagle's medium (Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies) and 1% penicillin–streptomycin (Life Technologies). The cells were maintained at 37 °C in a humidified chamber supplemented with 5% CO₂.

Mice for behavioral experiments

All methods were performed in accordance with the relevant guidelines and regulations. The *Frrs1l* heterozygous knockout mutant mice (C57BL/6N background) were purchased from the Medical Research Council (MRC; stock EM:07313). In this mutant mouse line, exon 3 of *Frrs1l* was deleted by insertion of the LacZ reporter gene, resulting in a frameshift. The knockout mice were maintained by breeding with wildtype C57BL/6N mice. All the mice were genotyped following the genotyping protocol of the MRC. Mice were assigned to experimental groups based on their genotypes. Adult (aged 2–3 months) *Frrs1l*^{+/-} and wildtype littermate mice of both genders were used for our behavioral experiments. The mice were group-housed (3–5 mice per cage). Constant temperature (22 ± 1 °C), humidity (40–60%), and 12 h light/dark cycle were maintained in a vivarium. Food and water were available ad libitum. The *Frrs1l*^{+/-} and littermate wildtype mice were habituated to handling for five days before the behavioral tests. Following the behavioral experiments, mice were euthanized using carbon dioxide (CO₂) inhalation. They were placed in a CO₂ chamber, and their respiration was monitored. After respiration ceased, mice remained in the chamber for an additional 60 s. To ensure the death of the animals, cervical dislocation was performed as a secondary method. All the procedures were performed in accordance with Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines and regulations, and were approved by the KAIST Institutional Animal Care and Use Committee.

For the generation of the humanized (*Frrs1l*^{h/m}) mice, Cyagen Biosciences Inc. (Santa Clara, US) generated the embryonic stem cells (ESCs) carrying the human allele by homologous recombination. Briefly, the targeting vector was designed to replace the coding sequence of mouse *Frrs1l* exon 1 with human *FRRS1L* cDNA. The targeting vector was electroporated into C57BL/6J ESCs and targeted ESC clones were identified using PCR and Southern blot. Chimeras were generated by microinjection of ESCs into blastocyst of albino C57BL/6J (Jackson Laboratory, stock #000058). Founders carrying the human allele were identified using genotyping and crossed with FLP deleters to remove the Neo cassette. Genotyping for the knockin and wildtype alleles was performed using primers that produced 587 bp and 510 bp bands, respectively (Supplementary Table 5).

We checked the expression of the human allele in the *Frrs1l*^{h/m} mice. The mice were anesthetized with isoflurane. Brains were extracted and ice-chilled in artificial cerebrospinal fluid. Total RNA was isolated from the whole brain with mirVANA miRNA isolation kit (AM1561, Ambion), according to the manufacturer's instruction. The cDNA was reverse-transcribed using miScript II RT kit (218160, Qiagen). The qRT-PCR was performed using QuantiTect SYBR Green PCR kit (204143, Qiagen) with a Rotor-Gene Q (Qiagen). Specific primer pairs designed to target *Frrs1l*^h and *Gapdh* are listed in Supplementary Table 6. The *Frrs1l*^h mRNA expression level was normalized using the *Gapdh* mRNA expression level by calculating the differences between their CT values represented as ΔCT ³⁸.

Three-chamber tests

The three-chamber test was performed following the procedures that were previously described³⁹ with minor modifications. The three-chamber apparatus comprised three identical acrylic chambers (20 cm × 40 cm × 25 cm; each, 60 cm × 40 cm × 25 cm). The test consisted of three sessions: habituation, sociability, and social novelty recognition. For the habituation, the subject mouse was placed in the center chamber and allowed to freely explore all three chambers and two identical empty wire cups (7 cm diameter, 15 cm height with 0.7 cm spaced grid bar) located at the center of the side chambers for 5 min. In the sociability test, the subject mouse was gently guided to the center chamber, and an age- and sex-matched conspecific stranger mouse was introduced into the wire cup of one of the side chambers. Designation of the side chambers was counterbalanced. The subject mouse was then allowed to explore all three chambers again for 10 min. In the social novelty recognition test, an age- and sex-matched novel conspecific mouse was introduced into the other wire cup, while the position of the familiar mouse was maintained. The subject mouse was allowed to explore for another 10 min and removed back to the home-cage. All stranger mice were habituated to the wire cup for 30 min before the test. Recorded videos were analyzed using EthoVision XT 10.1 software (Noldus). Direct contact of the subject mouse with the wire cup and stretching of the nose in an area 3 cm around the wire cup were counted as an interaction. The discrimination index (DI) was calculated as follows:

$$\text{Sociability DI} = \frac{\text{Interaction time (mouse)} - \text{Interaction time (empty cup)}}{\text{Interaction time (mouse)} + \text{Interaction time (empty cup)}} \times 100$$

$$\text{Social novelty recognition DI} = \frac{\text{Interaction time (novel mouse)} - \text{Interaction time (familiar mouse)}}{\text{Interaction time (novel mouse)} + \text{Interaction time (familiar mouse)}} \times 100$$

Social memory assessment

Two-trial social memory tests were performed as previously suggested⁴⁰. In the first trial, the subject mouse was placed in the test cage and allowed to explore for 15 min. After the habituation, a novel age-matched 129S6/C57BL/6J hybrid mouse (stimulus 1) was placed in the test cage together with the subject mouse for 2 min (first trial). After seven days, the subject mouse was re-exposed to the stimulus 1 mouse for 2 min (second trial). These procedures were used for the 7d memory test. Subsequently, the 1d memory test was performed two days later. The same subject mouse was exposed to a novel age-matched 129S6/C57BL/6J hybrid mouse (stimulus 2) in the first trial and re-exposed the next day to the stimulus 2 mouse in the second trial. Behaviors initiated by the subject mouse toward the stimulus mouse, including sniffing, licking, pawing, grooming, and close following, were measured as an investigation behavior. Mice with a duration of investigation < 24 s were excluded from the analysis⁴⁰.

Other behavior experiments

The open field test was performed in a white non-transparent acrylic open field chamber (40 cm × 40 cm × 40 cm). The subject mouse was placed in the wall-side of the chamber and recorded for 5 min. The center zone was defined as a 20 cm × 20 cm area centered on the chamber. The movement of the mouse was tracked using EthoVision XT 10.1 software. The total distance moved and time spent in the center zone were obtained as the indicators of locomotor activity and anxiety-like behavior, respectively.

The object location memory test was conducted following a previously reported protocol⁴¹ with minor modifications. Briefly, a white non-transparent acrylic-walled chamber (40 cm × 40 cm × 40 cm) with a vertical black marking strip was used. The subject mouse was habituated in a chamber with beddings for 5 min through six consecutive days. On the training day, two identical objects (cylinder shape, 7.5 cm diameter, and 4 cm height) were placed at pre-designated positions in the chamber. The mouse was allowed to freely explore the objects for 10 min. One of the objects was moved to a novel location 24 h later, and the retention test was conducted. The mouse was re-exposed to the chamber and allowed to explore the two objects each placed in the familiar or novel location for 10 min. Exploration time of a mouse was counted when the mouse's nose was within 1 cm of the object. Discrimination index was calculated as follows:

$$\frac{\text{Exploration time (novel object)} - \text{Exploration time (familiar object)}}{\text{Exploration time (novel object)} + \text{Exploration time (familiar object)}} \times 100$$

Contextual fear conditioning was performed in a conditioning chamber (Coulbourn Instruments). The subject mice were placed in the conditioning chamber for 2 min and allowed to explore it. Three electrical foot shocks (0.5 mA, 2 s) were delivered as an unconditioned stimulus via a metal grid with a 1 min inter-stimulus-interval. The mice were re-exposed to the conditioning chamber 24 h later and recorded for 5 min. Freezing behavior was defined as immobility of the subject mice (no movement except respiration). FreezeFrame software (Actimetrics) was used to automatically monitor the freezing behavior. All behavioral analyses were conducted in a blinded manner.

Prism6 (GraphPad software) was used for all statistical analyses of our behavioral experiments. Two-way analysis of variance followed by Bonferroni's post-hoc multiple comparisons test was performed to analyze the three-chamber interaction time results. An unpaired two-tail t-test was used for the analysis of the results from the three-chamber DI, open field test, object location memory test, and contextual fear conditioning.

Results

To detect SRVs from whole-genome sequences, we integrated 455 multiplex families from the MSSNG dataset with 469 multiplex families from the iHART dataset, excluding duplicate families from iHART. As a result, we analyzed the genome sequences of 1,955 children affected with ASD in 866 multiplex families (Fig. 1). Variants shared between siblings were identified, and we selected rare variants with minor allele frequencies lower than 0.01 in the general population. Thus, 17,122,961 variants were identified and termed SRVs. Recurrent SRVs were identified, and 4,371,196 SRVs were used in a downstream analysis. The majority of SRVs were located in the distal intergenic (~51%) and intron (~40%) regions (Supplementary Fig. 1). Among 18,712 genes mapped to the SRVs, we selected 874 that showed differential expression in ASD-affected brain tissue³⁰ and demonstrated a significant difference in accessible chromatin peaks between the human and chimpanzee brains⁴². To extract genes around each SRV that have been significantly affected by evolution, we calculated the evolutionary scores for each gene based on the frequency of the evolutionary sequence divergence sites (see the "Methods" section).

We then selected the top 100 genes for further examination (Fig. 2A). For prioritization, we calculated the frequency of the SRVs and measured the haplotype homozygosity^{22,24,25} for each gene. We found that four outlier genes, including *FRRS1L*, were highly ranked in both scoring schemes (Fig. 2B). We compared the gene expression in human tissues for the outlier genes using the GTEx RNA-seq data. The *FRRS1L* gene was specifically expressed in the brain (Fig. 2C). Particularly interesting is the lower expression of this gene in the ASD-affected brain compared to the normal brain³⁰ ($P=0.0076$), and in the macaque brain regions during early and midfetal ages compared to that of the humans³¹ (Supplementary Fig. 2). This gene is expressed in excitatory neurons and coexpressed with glutamate receptor signaling genes in fetal brains (Supplementary Figs. 3, 4).

In the genebody of the *FRRS1L* promoter, we found 63 different SRVs (SRVs track, Fig. 3). Among those, two particular SRVs were flanked by a human-specific fixation sequence within 10 bp (human-specific fixed

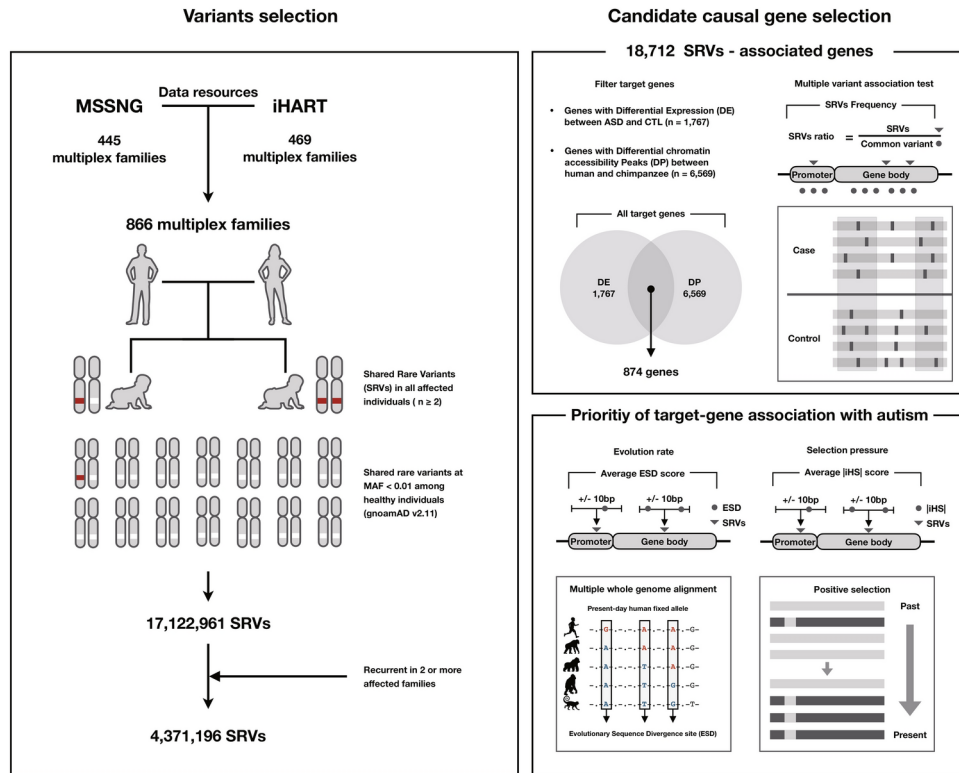


Fig. 1. Overview of the data analyses of ASD-affected families. Summary of the SRV identification (left) and gene selection processes (right) in 866 multiplex families. ASD, autism spectrum disorder; SRV, shared rare variant.

sequence track, Fig. 3). We tested whether those two fixation events were accompanied by positive selection signatures. Both the Tajima's D and nucleotide diversity (π) pinpointed a region (blue shade, Fig. 3) harboring a fixed CCC trinucleotide and C-to-T substitution in FAM_AU031404(MSSNG) and AU1651(iHART) within 6 bp from each other (red stars, Fig. 3). Such a variant, also found in the macaque genome, increases the binding affinity of the ZIC5 repressor, while disrupting the motif of the TFAP2A activator (Fig. 4A and Supplementary Fig. 5A,B).

To support our finding, we performed a CRISPR/Cas9 experiment to introduce the ancestral nucleotides. The results demonstrated an expected transcriptional consequence (Supplementary Fig. 6A), which indicated gene downregulation by C-to-T substitution (Fig. 4B,C). The TFAP2A motif was disrupted by not only the ASD/macaque variant but also the ancestral sequences at the human-specific trinucleotide (Fig. 4A and Supplementary Fig. 5C). In other words, the motif is most active in wild-type humans. The introduction of the ancestral nucleotides by CRISPR/Cas9 (Supplementary Fig. 6B) induced gene downregulation (Fig. 4B,C).

A more recent selection can be tested with segregating polymorphisms. We examined the measure of the extended haplotype homozygosity^{22,24,25} in the human population across the gene. High levels of the iHS²⁴ were observed near the promoter (Fig. 3). Two SNPs (rs2176692 and rs75742991) in close proximity to the SRVs (green shade, Fig. 3 and Supplementary Fig. 7) were significant eQTL for *FRRS1L* (Fig. 4D and Supplementary Fig. 7). Importantly, the derived alleles, significantly increasing the expression level of *FRRS1L*, were on the selected haplotypes (Fig. 4E and Supplementary Fig. 7). Another line of evidence for the selective advantage of *FRRS1L* upregulation came from a large-scale GWAS of educational attainment²⁹. The effective allele for cognitive performance at rs2439649 was associated with gene activation and positive selection (Supplementary Fig. 8).

The ASD-affected children of AU031404 scored extremely low in the domain of socialization, according to the Vineland Adaptive Behavior Scales (Supplementary Fig. 9). Thus, we examined social behaviors of the heterozygous knockout (*Frrs1l*^{+/-}) mice by performing the consecutive three-chamber tests for sociability and social novelty recognition. In the sociability test (Fig. 5A), both *Frrs1l*^{+/-} and littermate wildtype mice preferred the interaction with a conspecific mouse, indicating normal sociability for the *Frrs1l*^{+/-} mice (Fig. 5A–G and Supplementary Movies 1, 2). In the following test, a novel mouse was introduced into the empty chamber (Fig. 5H). As predicted, the wildtype mice showed a preference for a novel mouse over a familiar one. In contrast, the *Frrs1l*^{+/-} mice exhibited no preference for social novelty (Fig. 5I and Supplementary Movies 3, 4). The discrimination index of the *Frrs1l*^{+/-} mice was significantly lower than that of the wild-type mice (Fig. 5J). There was no significant difference in total interaction time (Mouse + Empty or Familiar + Novel) between groups (Fig. 5K), suggesting that WT and *Frrs1l*^{+/-} mice did not differ in social interest or motivation for social

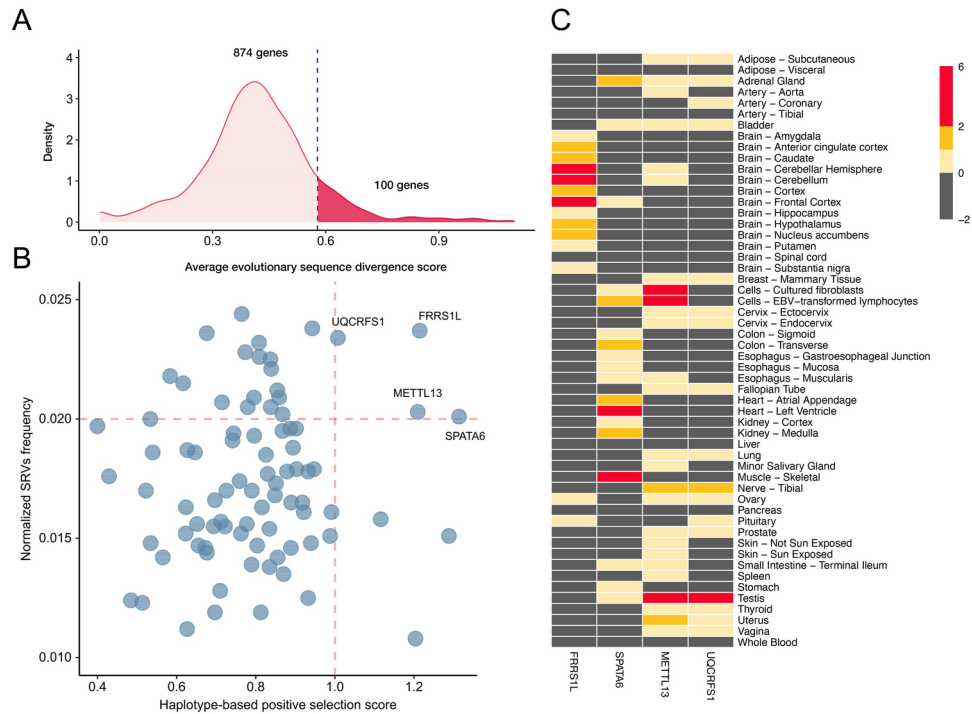


Fig. 2. Selection of target genes associated with ASD. **(A)** Density plot for the evolutionary scores. The average number of evolutionary sequence divergence sites around the SRVs is marked for each gene. **(B)** Gene prioritization by evolutionary scores. Top 100 genes selected from **(A)** are plotted, according to the degree of normalized SRV's frequency and haplotype homozygosity. **(C)** Tissue specificity for the expression of target genes. Gene expression heatmaps are constructed with GTEx v8 (54 human tissue type). Gene expression values represent the Z-score normalized across the tissue type per gene. ASD, autism spectrum disorder; SRV, shared rare variant.

engagement. In addition, there was no significant difference in time spent in center chamber between groups (Fig. 5L).

To further confirm the discrimination index results, we analyzed another index of social behaviors, 'the number of interactions' and found consistent results. The number of interactions was higher to the mouse than to the empty in both WT and *Frrs1l*^{+/-} mice with no significant group difference (Fig. 5F). However, in the social novelty recognition test, while WT mice showed a higher number of interactions to the novel mouse than to the familiar mouse, *Frrs1l*^{+/-} mice showed a similar number of interactions to both the novel and familiar mouse (Fig. 5M). Notably, the number of interactions to the familiar mice was increased in HT mice compared with WT. We also analyzed the 'mean duration of single interaction time' (mean duration of time spent during one bout of interaction) and found no significant differences in all comparisons performed. (Fig. 5N). These results suggest that the primary source of variation in the interaction time observed is the number of interactions but not the duration of single interaction time. Together, these results revealed a significant deficit in social novelty recognition in the *Frrs1l*^{+/-} mice.

In the open field tests, the *Frrs1l*^{+/-} and wildtype mice showed comparable levels of moving distance, indicating normal locomotor activity (Supplementary Fig. 10A,B). The thigmotaxis indices were also similar, suggesting no significant change in basal anxiety level in *Frrs1l*^{+/-} mice compared to WT mice (Supplementary Fig. 10C–G). To test whether other anxiety functions are normal in the *Frrs1l*^{+/-} mice, we assessed the hippocampus-dependent memory tasks. There was no significant difference between the *Frrs1l*^{+/-} and wildtype mice in the test for object location memory (Supplementary Fig. 10H,I). The *Frrs1l*^{+/-} and wildtype mice exhibited similar levels of freezing in the contextual fear memory test (Supplementary Fig. 10J). Therefore, our results showed that the *Frrs1l*^{+/-} mice have specific impairment in social novelty recognition, while they are normal for other behavioral characteristics.

We noted a substantial difference between the human and mouse sequences of the gene (Fig. 6). This prompted us to test the functional differences between the two homologs. To this end, we generated humanized mice, in which one of the mouse copies of the *FRRS1L* gene was replaced with the human sequence (*Frrs1l*^{h/m}). Quantitative RT-PCR analysis confirmed the expression of the human allele in *Frrs1l*^{h/m} mice (Supplementary Fig. 11). However, the three-chamber tests did not show significant differences in sociability and social novelty recognition (Supplementary Fig. 12). To further address the social memory function of the *Frrs1l*^{h/m} mice, we used a two-trial social memory test that assesses an animal's ability to retain social memory over time. In this test, the investigation duration is compared between the first and second exposure with either a one-day or seven-day interval (1d and 7d memory test, respectively). If the social memory is retained, the mice would spend

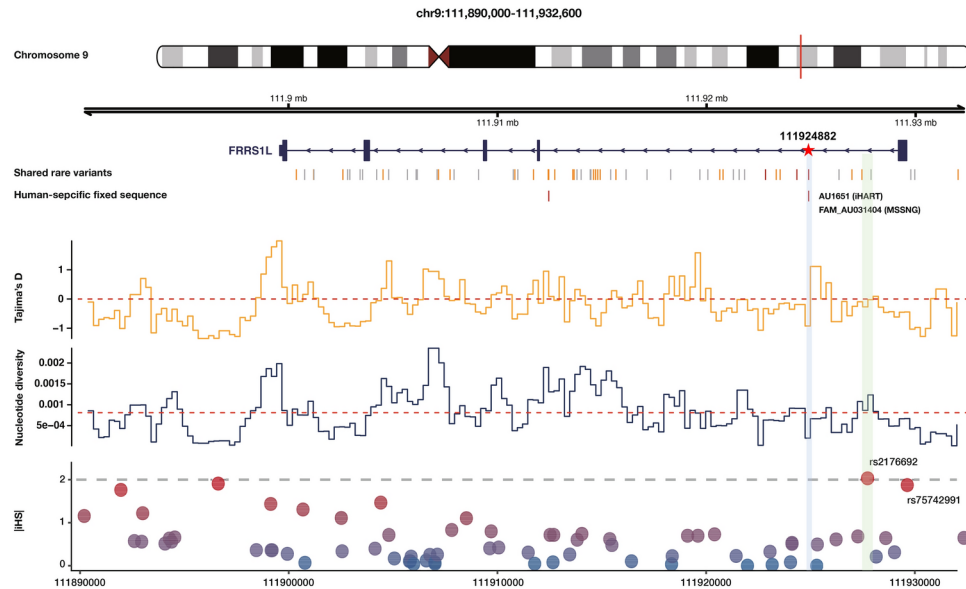


Fig. 3. Evolutionary analysis of the *FRRS1L* region. Evolutionary genetic analysis for 63 SRVs in *FRRS1L*. In the shared rare variants track below, each bar indicates 63 SRVs, and the different colors indicate the number of evolutionary sequence divergence sites at ± 10 bp windows around the SRVs (gray = 0, orange = 1, and red > 2). In the human-specific fixed sequence track below, the red bar indicates the position of the human-specific sequence at ± 10 bp windows around the SRVs. In the iHS track below, the single-nucleotide polymorphism with $|iHS| > 2$ (rs2176692) is highlighted in red. Tajima's D and nucleotide diversity (π) were obtained using 1-kb sliding windows. The dots indicate the values corresponding to the fixed sites with the one indicating positive selection highlighted in red. SRV, shared rare variant; iHS, integrated Haplotype Score.

less time interacting with the stimulus mouse during the second exposure compared to the first exposure. This two-trial social memory test effectively detects long-term social memory lasting 1–3 days in C57BL/6 J mice. However, this social recognition is no longer retained after 7 days interval⁴⁰. We hypothesized that if *Frrs1l*^{h/m} mice has an enhanced social memory, the mice might show memory retention even in 7d test. To test this hypothesis, we performed the 7d memory test first and then the 1d test using two different stimulus mice for the same subject mouse (Fig. 7A). In the 7-day memory test, *Frrs1l*^{h/m} mice demonstrated a significantly reduced investigation duration between the first and second exposures, whereas the wild-type mice did not (Fig. 7B,C). These data indicate that only the humanized mice retained social memory for seven days. According to the 1d memory test, both WT and *Frrs1l*^{h/m} subject mice were normal in 1-day social memory (Fig. 7D,E). Together, these results suggest that the humanized mice exhibit enhanced social memory retention.

Discussion

Our population genetic analyses identified multiple variants with signs of positive selection in proximity to the ASD family variants. One of them was a fixed variant residing within a binding site shared by two related TFs. In addition to this old variant, there were segregating polymorphisms indicating a more recent selection. We validated the functionality of these variants by leveraging the CRISPR/Cas9 and eQTL analyses. Our results propose a molecular mechanism that may contribute to survival and reproduction conferred by social cognition and long-term social memory^{43–46}.

We identified risk genes in multiplex families by concentrating on rare variants. Compared to simplex families, transmitted private variants play a more prominent role in multiplex families⁴⁷. Therefore, genes that harbor rare variants that are shared between siblings with ASD may be distinct from known ASD risk genes that harbor de novo variants. However, the SRVs identified in this study are likely to have lower penetrance, as well as lower effect size, necessitating prioritization by evolutionary signals. Through the prioritization process, a candidate gene, *FRRS1L*, was identified.

The *FRRS1L* gene is known to be an outer core component of the AMPA receptor complex, which mediates the glutamatergic excitatory signals^{48,49}. A *FRRS1L* knockdown was shown to reduce AMPA-mediated currents in human neural cells⁵⁰. Clinically, *FRRS1L* is associated with developmental and epileptic encephalopathy-37 (DEE-37, OMIM#616,981), which show autosomal recessive pattern of inheritance^{50,51}. In previous studies, children with homozygous loss-of-function mutations in *FRRS1L* exhibited seizure onset between 6 and 30 months of age, followed by regression in motor and speech development. Additionally, hypotonia and chorea was observed, and brain MRI in some children revealed progressive cortical and cerebellar atrophy. In an animal model, mice lacking *Frrs1l* presented with early onset motor deficits, hyperactivity, working memory impairments, and sleep fragmentation⁵². Abnormal electroencephalographic changes were noted. The loss of *Frrs1l* resulted in increased proportion of immature AMPA receptors, and decreased level of functional AMPA receptors at neuronal synapses. To our knowledge, an association between *FRRS1L* and ASD has not yet been

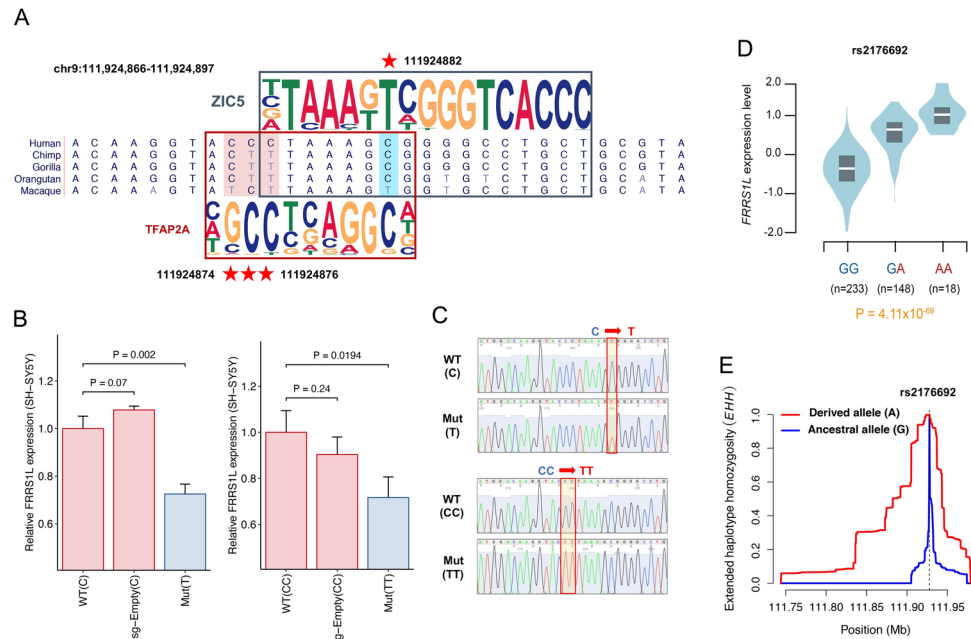


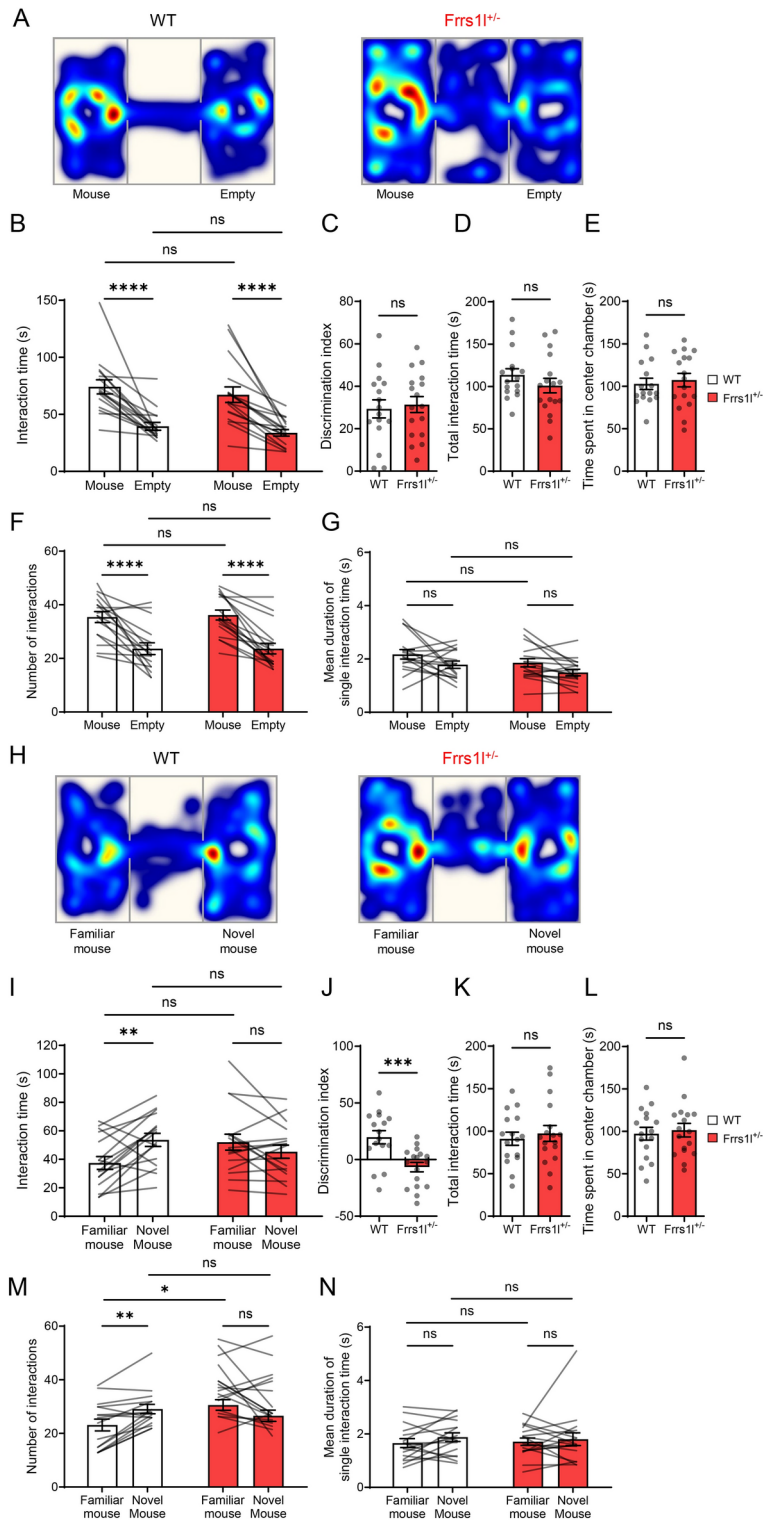
Fig. 4. Functional analysis of the *FRRS1L* regulatory variants. **(A)** Changes in transcription factor binding motifs by the fixed trinucleotide with positive selection signatures and its flanking SRV. Calculation of the affinity changes is provided in Supplementary Fig. 5. **(B)** Expression level of *FRRS1L* measured by qRT-PCR in SH-SY5Y cells before (red) and after (blue) CRISPR/Cas9-mediated genome editing. Relative expression levels were computed by dividing with the wildtype measure. Three technical replicates were obtained for each of the three independent experiments. Shown data is a representative of the three experiments. P values were derived from two-tailed Student's t-tests. Error bars, SEM. **(C)** Confirmation of CRISPR-induced edits with Sanger sequencing chromatographs for the autism/maaque variant (upper) and the ancestral sequences at the human-specific fixation site (lower). **(B–C)** WT, wildtype sequence; sg-Empty, no sgRNA; Mut, intended mutation. **(D)** eQTL mapping results for the rs2176692 genotypes and normalized *FRRS1L* expression levels. **(E)** Decay of the EHH statistic²², as a function of distance for haplotypes that carry the derived and ancestral allele of rs2176692. qRT-PCR, quantitative real-time reverse-transcription; SEM, standard error of mean; EHH, extended haplotype homozygosity.

reported. However, AMPA receptors are known to play an integral role in long-term potentiation, which is the cellular basis of learning and memory^{53–55}. Dysfunction of the AMPA receptors had been implicated in ASD^{56–61}.

In the gene body and promoter regions of *FRRS1L*, we searched for genomic regions that have become fixed in humans. By doing so, we aimed to find SRVs that altered human-specific fixed sequences, thereby have the potential to affect human-specific behaviors, such as social behaviors. Of the two SRVs located in the human-specific fixed sequence, one altered the binding affinity of two different TFs, ZIC5 and TFAP2A. Notably, the two TFs were closely related in that they are both expressed in neural crest cells and that their deficiency results in neural tube defects^{62,63}. When the detected SRV, which is also an ancestral allele found in macaques, was introduced in vitro by the CRISPR/Cas9 experiment, *FRRS1L* expression was decreased, suggesting SRV functionality. Human-specific fixed sequence, also known as the human accelerated regions, are enriched within 1 Mbp of a gene, implying that they play a role in the transcriptional process⁶⁴. The association between these accelerated regions and neurodevelopmental disorders, such as schizophrenia, has been reported before. Specifically, GABAergic and glutamatergic genes stood out among the related genes⁶⁵. Yet in another study, genes related to human accelerated regions were found to be associated with ASD. Taken together, our results suggest that the SRV found in this study may influence human-specific social behavior by changing the expression level of the *FRRS1L* gene.

Unlike other mouse models of ASD^{66–69}, *Frrs1l*^{+/−} specifically affected social novelty recognition without altering other behavioral phenotypes. However, in *Frrs1l* null mutant mice (*Frrs1l*^{−/−} mice), Stewart et al.⁵² reported neurological abnormalities, with no available information on deficits in sociability or social novelty recognition. Given alterations in the motor functions, it might not be feasible to properly test the social behaviors in the *Frrs1l* null mutant mice. In our study, we found no noticeable alterations in motor functions as demonstrated in the open field test. One possible interpretation of the results from the two knockout mouse models is there might be a gene-dosage effect on the abnormalities in neurological and cognitive phenotypes associated with the loss of FRRS1L functions such that motor impairment is elicited only in the null mutant condition, whereas some aspects of social behaviors such as social novelty recognition are more sensitive to the loss of FRRS1L functions.

Mice with the human sequence, *Frrs1l*^{h/m}, showed retention of long term (7 days) social memory, in contrast to wild-type mice, which demonstrated a lack of long-term social memory retention over the same period.



Considering that social memory formation and retention can be influenced by various motor, sensory, and cognitive functions, further research employing diverse behavioral models may elucidate the specific relationship between the FRRS1L sequence and social memory.

Limitations

This study has some limitations. First, although widely used, animal models are approximations of human behavior. Therefore, findings from our mouse experiments may not be generalized to humans. It is important to acknowledge ongoing debates regarding the limitations of the three-chamber test⁷⁰. Given the complexity of social behaviors in natural settings, the three-chamber test captures a limited range of variables, such as the time spent interacting with a stimulus mouse. Since the stimulus mouse is confined within a wire cup, the molecular mechanisms underlying social interactions in this test may differ from those involved in direct

◀ **Fig. 5.** Social behaviors of the *Frrs11*^{+/-} mice. (A–E) Results of the sociability tests. (A) Representative heat map images. (B) Both the wildtype and *Frrs11*^{+/-} mice preferred interaction with a conspecific mouse to the empty wire cup (n = 16, WT; n = 17, *Frrs11*^{+/-}; stimulus × genotype *F* (1,31) = 0.0178, *P* = 0.8949; stimulus *F* (1,31) = 57.11, *P* < 0.0001; genotype *F* (1,31) = 1.236, *P* = 0.2749; Two-way repeated measures ANOVA with Bonferroni’s multiple comparisons test). (C) There was no significant difference in the discrimination index (*t* (31) = 0.3550, *P* = 0.725, unpaired two-tailed t-test). (D) There was no significant difference in the total interaction time, defined as the sum of the durations during which the subject mice interacted with the conspecific mouse and the empty cup (*t* (31) = 1.112, *P* = 0.2749, unpaired two-tailed t-test). (E) There was no significant difference in time spent in center chamber during sociability tests (*t* (31) = 0.4247, *P* = 0.6740, unpaired two-tailed t-test). (F) Results of number of interactions in the sociability tests (stimulus × genotype *F* (1,31) = 0.0502, *P* = 0.8243; stimulus *F* (1,31) = 57.82, *P* < 0.0001; genotype *F* (1,31) = 0.0257, *P* = 0.8736; Two-way repeated measures ANOVA with Bonferroni’s multiple comparisons test). (G) Results of mean duration of single interaction time (mean duration of time spent during one bout of interaction) in the sociability tests (stimulus × genotype *F* (1,31) = 0.0072, *P* = 0.9329; stimulus *F* (1,31) = 7.961, *P* = 0.0083; genotype *F* (1,31) = 3.547, *P* = 0.0691; Two-way repeated measures ANOVA with Bonferroni’s multiple comparisons test). (H–N) Results of the tests for social novelty recognition tests. (H) Representative heat map images. (I) Whereas the wildtype mice spent more time interacting with a novel mouse than with a familiar mouse, the *Frrs11*^{+/-} mice showed no significant difference in preference to the familiar and novel mouse, with a trend toward slightly increased preference to the familiar mouse compared to WT mice (stimulus × genotype *F* (1,31) = 12.83, *P* = 0.0012; stimulus *F* (1,31) = 2.223, *P* = 0.1461; genotype *F* (1,31) = 0.2662, *P* = 0.6095; Two-way repeated measures ANOVA with Bonferroni’s multiple comparisons test). (J) Discrimination index of the *Frrs11*^{+/-} mice was significantly lower than that of the wildtype mice (*t* (31) = 3.733, *P* = 0.0008, unpaired two-tailed t-test). (K) There was no significant difference in the total interaction time, defined as the sum of the durations during which the subject mice interacted with the familiar mouse and the novel mouse (*t* (31) = 0.5160, *P* = 0.6095, unpaired two-tailed t-test). (L) There was no significant difference in time spent in center chamber during social novelty recognition tests (*t* (31) = 0.3744, *P* = 0.7107, unpaired two-tailed t-test). (M) Results of number of interactions in the social novelty recognition tests (stimulus × genotype *F* (1,31) = 15.14, *P* = 0.0005; stimulus *F* (1,31) = 0.5754, *P* = 0.4538; genotype *F* (1,31) = 0.9568, *P* = 0.3356; Two-way repeated measures ANOVA with Bonferroni’s multiple comparisons test). (N) Results of mean duration of single interaction time in the social novelty recognition tests (stimulus × genotype *F* (1,31) = 0.2341, *P* = 0.6319; stimulus *F* (1,31) = 1.398, *P* = 0.2461; genotype *F* (1,31) = 0.0029, *P* = 0.9572; Two-way repeated measures ANOVA with Bonferroni’s multiple comparisons test). Data are presented as mean ± s.e.m. ns, not significant. **P* < 0.05, ***P* < 0.01, ****P* < 0.0001.

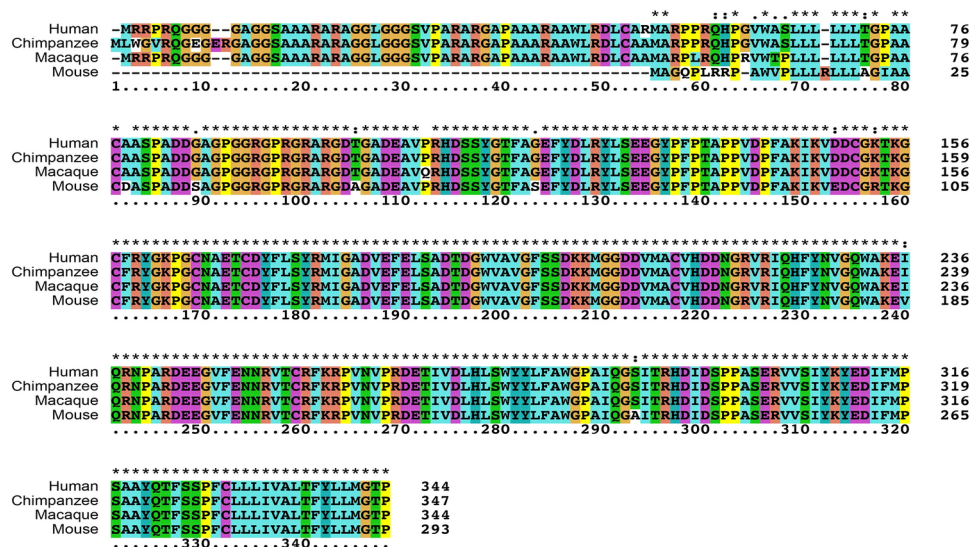


Fig. 6. Comparative protein sequence of *FRRS1L* in human, nonhuman primates, and mouse. Sequence alignments of *FRRS1L* orthologs. The following Ensembl entries were used: human (ENSP00000477141.1), chimpanzee (ENSPTRP00000036282.4), macaque (ENSMMUP00000023050.3), and mouse (ENSMUSP00000052507.5).

social interactions⁷¹. In future research, assessing social behavior in a more naturalistic environment would enhance the ecological validity of the findings⁷². Moreover, while age- and sex- matched stimulus mice were employed in three-chamber tests and social memory assessments, the body weights were not controlled for. Although age- and sex-matched mice typically have similar body weights in our experience, we cannot rule out

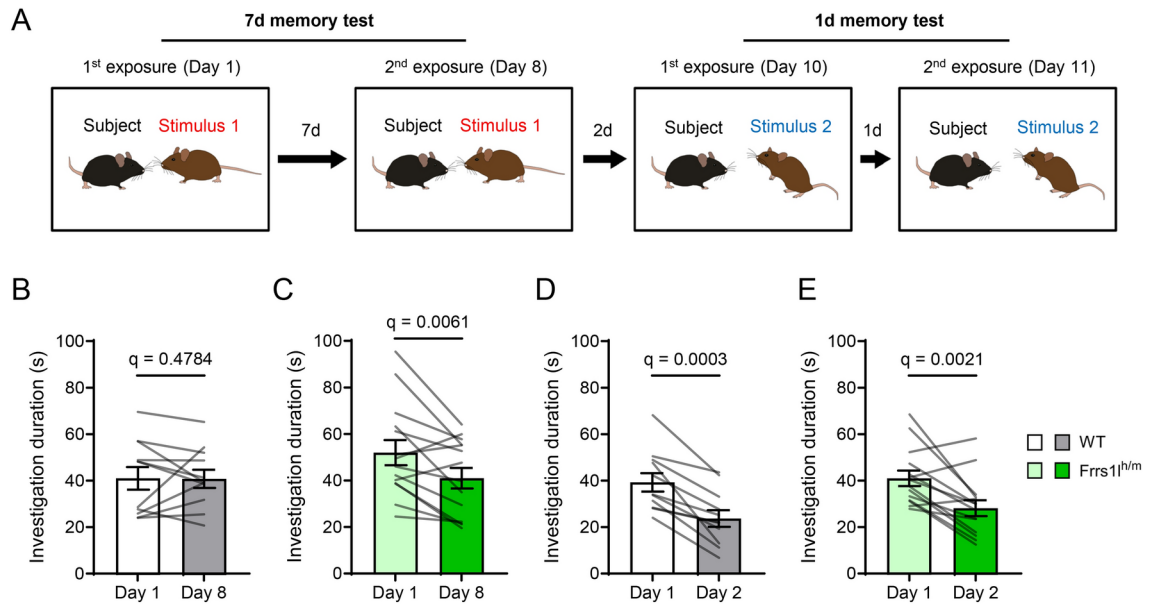


Fig. 7. Improved retention of social memory in the *Frrs1^{h/m}* mice. **(A)** Experimental procedures of the social memory assessments. **(B,C)** Results of the 7d memory tests. **(B)** Wildtype mice exhibited comparable investigation duration between the first and second exposure to the stimulus mouse ($n = 11$, WT; $n = 14$, *Frrs1^{h/m}*; $t(10) = 0.0677$, $P = 0.9474$, $q = 0.4784$, multiple paired t-tests with FDR correction). **(C)** *Frrs1^{h/m}* mice exhibited reduced investigation duration during the second exposure compared to the first exposure ($t(13) = 3.066$, $P = 0.0090$, $q = 0.0061$, multiple paired t-tests with FDR correction). **(D,E)** Results of the 1d memory tests. Both wildtype and *Frrs1^{h/m}* mice exhibited reduced investigation duration during the second exposure compared to the first exposure: **(D)** WT mice ($t(10) = 5.810$, $P = 0.0002$, $q = 0.0003$, multiple paired t-tests with FDR correction), **(E)** *Frrs1^{h/m}* mice ($t(13) = 3.832$, $P = 0.0021$, $q = 0.0021$, multiple paired t-tests with FDR correction). q -value is FDR-adjusted P -value. Data are presented as mean \pm s.e.m.

the possibility that differences in body weight may have introduced a confounding variable in the assessment of social behaviors in this study.

Second, to our knowledge, *FRRS1L* has not previously been reported as an ASD risk gene, including in the SFARI database. This might be because previous studies on ASD risk genes have primarily focused on rare coding variants, whereas our study emphasizes evolutionary signals of SRVs in noncoding regions. However, the results of this study warrant further validation in future studies. Third, we used *in silico* tools to predict the off-target effects of CRISPR/Cas9 experiments. Since experimental validation was not performed, this represents a limitation of our study. Finally, the molecular mechanism linking *FRRS1L* and altered social behavior warrants further investigation.

Conclusion

Based on the population genetic and evolutionary analyses in our study, we present SRVs associated with ASD. The CRISPR/Cas9 experiment introducing a single nucleotide change demonstrated that the found SRVs affect the expression levels of the candidate gene, *FRRS1L*. Furthermore, mouse behavioral experiments indicated that the gene is linked to social novelty recognition, suggesting a potential role for *FRRS1L* in ASD-related behaviors.

Data availability

Data and materials for this study can be found in the Supplementary information. Additional data and materials associated with this study are available from the corresponding author, Jung Kyoong Choi, upon reasonable request. Researchers can obtain the dataset used in this study by applying at the MSSNG (<https://research.mss.ng/>) and the iHART (<http://www.ihart.org>) online portal.

Received: 2 October 2023; Accepted: 17 December 2024

Published online: 03 January 2025

References

1. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: On the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).
2. Baio, J. et al. Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. *MMWR Surveill. Summaries.* **67** (2018).
3. Gaugler, T. et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
4. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron.* **87**, 1215–1233 (2015).

5. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. **515**, 216–221 (2014).
6. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. **515**, 209–215 (2014).
7. Turner, T. N. et al. Genomic patterns of de novo mutation in simplex autism. *Cell*. **171**, 710–722.e12 (2017).
8. C Yuen, R. K. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
9. Kaplanis, J. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*. **586**, 757–762 (2020).
10. Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*. **180**, 568–584.e23 (2020).
11. Ruzzo, E. K. et al. Inherited and de novo genetic risk for autism impacts shared networks. *Cell*. **178**, 850–866.e26 (2019).
12. Turner, T. N. et al. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am. J. Hum. Genet.* **98**, 58–74 (2016).
13. Kim, I. B. et al. Non-coding de novo mutations in chromatin interactions are implicated in autism spectrum disorder. *Mol. Psychiatry*. **27**, 4680–4694 (2022).
14. Williams, S. M. et al. An integrative analysis of non-coding regulatory DNA variations associated with autism spectrum disorder. *Mol. Psychiatry*. <https://doi.org/10.1038/s41380-018-0049-x> (2018).
15. Doan, R. N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell*. **167**, 341–354.e12 (2016).
16. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science*. **314**, 786 (2006).
17. Pollard, K. S. et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168 (2006).
18. Brandler, W. M. et al. Paternally inherited cis-regulatory structural variants are associated with autism. *Science*. **360**, 327–331 (2018).
19. Yuen, R. K. C. et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* **21**, 185–191 (2015).
20. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. **491**, 56–65 (2012).
21. Biswas, S. & Akey, J. M. Genomic insights into positive selection. *Trends Genet.* **22**, 437–446 (2006).
22. Sabeti, P. C. et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. **419**, 832–837 (2002).
23. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. **123**, 585–595 (1989).
24. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
25. Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. **449**, 913–918 (2007).
26. Cadzow, M. et al. A bioinformatics workflow for detecting signatures of selection in genomic data. *Front. Genet.* **5**, 293 (2014).
27. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
28. Yu, G., Wang, L. G. & He, Q. Y. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. **31**, 2382–2383 (2015).
29. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
30. Parikhshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. **540**, 423–427 (2016).
31. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science*. **362**, eaat8077 (2018).
32. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. **362**, eaat7615 (2018).
33. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
34. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature*. **508**, 199–206 (2014).
35. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*. **533**, 125–129 (2016).
36. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*. **30**, 1473–1475 (2014).
37. Park, J., Bae, S. & Kim, J. S. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics*. **31**, 4014–4016 (2015).
38. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods*. **25**, 402–408 (2001).
39. Kaidanovich-Beilin, O., Lipina, T., Vukobradovic, I., Roder, J. & Woodgett, J. R. Assessment of social interaction behaviors. *J. Vis. Exp.* **48**, 2473 (2011).
40. Gao, X.-M., Elmer, G. I., Adams-Huet, B. & Tamminga, C. A. Social memory in mice: disruption with an NMDA antagonist and attenuation with antipsychotic drugs. *Pharmacol. Biochem. Behav.* **92**, 236–242 (2009).
41. Vogel-Ciernia, A. & Wood, M. A. Examining object location and object recognition memory in mice. In *Curr Protoc Neurosci*. 8.31.1–8.31.17 (Wiley, 2014).
42. Kanton, S. et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* (2019).
43. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science*. **211**, 1390–1396 (1981).
44. McComb, K., Moss, C., Durant, S. M., Baker, L. & Sayialel, S. Matriarchs as repositories of social knowledge in African elephants. *Science*. **292**, 491–494 (2001).
45. Boeckle, M. & Bugnyar, T. Long-term memory for affiliates in ravens. *Curr. Biol.* **22**, 801–806 (2012).
46. Bruck, J. N. Decades-long social memory in bottlenose dolphins. *Proc. R. Soc. B Biol. Sci.* **280**, 20131726 (2013).
47. Wilfert, A. B. et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat. Genet.* **53**, 1125–1134 (2021).
48. Schwenk, J. et al. High-resolution proteomics unravel architecture and molecular diversity of native AMPA receptor complexes. *Neuron*. **74**, 621–633 (2012).
49. Schwenk, J. et al. Regional diversity and developmental dynamics of the AMPA-receptor proteome in the mammalian brain. *Neuron*. **84**, 41–54 (2014).
50. Madeo, M. et al. Loss-of-function mutations in FRRS1L lead to an epileptic-dyskinetic encephalopathy. *Am. J. Hum. Genet.* **98**, 1249–1255. <https://doi.org/10.1016/j.ajhg.2016.04.008> (2016).
51. Shaheen, R., Al Tala, S., Ewida, N., Abouelhoda, M. & Alkuray, F. S. Epileptic encephalopathy with continuous spike-and-wave during sleep maps to a homozygous truncating mutation in AMPA receptor component FRRS1L. *Clin. Genet.* **90**, 282–283 (2016).
52. Stewart, M. et al. Loss of Frrs1l disrupts synaptic AMPA receptor function, and results in neurodevelopmental, motor, cognitive and electrographical abnormalities. **44**, 1–36 (2018).
53. Rumpel, S. Postsynaptic receptor trafficking underlying a form of associative learning. *Science*. **308**, 83–88 (2005).
54. Derkach, V. A., Oh, M. C., Guire, E. S. & Soderling, T. R. Regulatory mechanisms of AMPA receptors in synaptic plasticity. *Nat. Rev. Neurosci.* **8**, 101–113 (2007).
55. Kessels, H. W. & Malinow, R. Synaptic AMPA receptor plasticity and behavior. *Neuron*. **61**, 340–350 (2009).

56. Salpietro, V. et al. AMPA receptor GluA2 subunit defects are a cause of neurodevelopmental disorders. *Nat. Commun.* **10**, 3094 (2019).
57. Chen, J., Yu, S., Fu, Y. & Li, X. Synaptic proteins and receptors defects in autism spectrum disorders. *Front. Cell Neurosci.* **8**, 1–13 (2014).
58. Etherton, M. R., Tabuchi, K., Sharma, M., Ko, J. & Südhof, T. C. An autism-associated point mutation in the neuroligin cytoplasmic tail selectively impairs AMPA receptor-mediated synaptic transmission in hippocampus. *EMBO J.* **30**, 2908–2919 (2011).
59. Bozdagi, O. et al. Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. Autism.* **1**, 15 (2010).
60. Ramanathan, S. et al. A case of autism with an interstitial deletion on 4q leading to hemizyosity for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. *BMC Med. Genet.* **5**, 10 (2004).
61. Purcell, A. E., Jeon, O. H., Zimmerman, A. W., Blue, M. E. & Pevsner, J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. *Neurology.* **57**, 1618–1628 (2001).
62. Zhang, J. et al. Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. *Nature.* **381**, 238–241 (1996).
63. Inoue, T. et al. Mouse *Zic5* deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives. *Dev. Biol.* **270**, 146–162 (2004).
64. Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. R. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20130025 (2013).
65. Xu, K., Schadt, E. E., Pollard, K. S., Roussos, P. & Dudley, J. T. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol. Biol. Evol.* **32**, 1148–1160 (2015).
66. DeLorey, T. M., Sahbaie, P., Hashemi, E., Homanics, G. E. & Clark, J. D. Gabrb3 gene deficient mice exhibit impaired social and exploratory behaviors, deficits in non-selective attention and hypoplasia of cerebellar vermal lobules: A potential model of autism spectrum disorder. *Behav. Brain Res.* **187**, 207–220 (2008).
67. Radyushkin, K. et al. Neuroligin-3-deficient mice: Model of a monogenic heritable form of autism with an olfactory deficit. *Genes Brain Behav.* **8**, 416–425 (2009).
68. Han, S. et al. Autistic-like behaviour in *Scn1a*^{+/-} mice and rescue by enhanced GABA-mediated neurotransmission. *Nature.* **489**, 385–390 (2012).
69. Pisansky, M. T. et al. Mice lacking the chromodomain helicase DNA-binding 5 chromatin remodeler display autism-like characteristics. *Transl. Psychiatry.* **7**, e1152 (2017).
70. Jabarin, R., Netser, S. & Wagner, S. Beyond the three-chamber test: toward a multimodal and objective assessment of social behavior in rodents. *Mol Autism.* <https://doi.org/10.1186/s13229-022-00521-6> (2022).
71. Fukumitsu, K. et al. Amylin-Calcitonin receptor signaling in the medial preoptic area mediates affiliative social behaviors in female mice. *Nat. Commun.* <https://doi.org/10.1038/s41467-022-28131-z> (2022).
72. Silverstein, S. M. et al. A comparative study of the MATRICS and IntegNeuro cognitive assessment batteries. *J. Clin. Exp. Neuropsychol.* **32**, 937–952 (2010).

Author contributions

KSL: Data curation, formal analysis, methodology, experiment, visualization, writing original draft; TL: Data curation, formal analysis, software, visualization, writing original draft, funding acquisition; MK: Methodology, experiment, visualization, writing original draft; EI: Investigation, methodology; H-JB: Investigation, methodology; MKS: Conceptualization, software; YK: Formal analysis; Y-JK: Conceptualization, experiment; J-HH: Conceptualization, experiment, reviewing and editing draft, funding acquisition; JKC: Conceptualization, reviewing and editing draft, funding acquisition, supervision. EI, H-JB: Investigation, methodology; MKS: Conceptualization, software; YK: Formal analysis; Y-JK: Conceptualization, experiment; J-HH: Conceptualization, experiment, reviewing and editing draft, funding acquisition; JKC: Conceptualization, reviewing and editing draft, funding acquisition, supervision.

Funding

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1067177) and a grant (2023IP0124) from the Asan Institute for Life Sciences, Asan Medical Center, Seoul, Korea.

Declarations

Ethics approval and consent to participate

All the procedures were performed in accordance with Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines and regulations, and were approved by the KAIST Institutional Animal Care and Use Committee.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83839-w>.

Correspondence and requests for materials should be addressed to J.-H.H. or J.K.C.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024