Rare variants in the outcome of social skills group training for autism

Danyang Li¹,², Nora Choque Olsson¹,³,⁴, Martin Becker¹,², Abishek Arora¹,², Hong Jiao⁵, Nina Norgren⁶, Ulf Jonsson¹,⁷, Sven Bölte¹,⁸*, Kristiina Tammimies¹,²*,#

¹Center of Neurodevelopmental Disorders (KIND), Centre for Psychiatry Research; Department of Women’s and Children’s Health, Karolinska Institutet and Child and Adolescent Psychiatry, Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden
²Astrid Lindgren Children’s Hospital, Karolinska University Hospital, Region Stockholm, Sweden
³Department of Psychology, Stockholm University, Stockholm, Sweden
⁴Center for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet and Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden
⁵Department of Biosciences and Nutrition, Karolinska Institutet, and Clinical Research Centre, Karolinska University Hospital, Huddinge, Sweden
⁶Department of Molecular Biology, National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Umeå University, 901 87 Umeå, Sweden
⁷Department of Neuroscience, Child and Adolescent Psychiatry, Uppsala University, Sweden
⁸Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, Western Australia

*equal contribution
#corresponding author

Information of the corresponding author:
Kristiina Tammimies
E-mail address: kristiina.tammimies@ki.se

Sven Bölte
E-mail address: sven.bolte@ki.se

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background

Exome sequencing has been proposed as the first-tier genetic testing in autism spectrum disorder (ASD). Here, we performed exome sequencing in autistic individuals with average to high intellectual abilities (N = 207) to identify a molecular diagnosis of ASD and genetic modulators of intervention outcomes following social skills group training (SSGT) or standard care.

Methods

Within a randomized controlled trial of SSGT, we performed exome sequencing to prioritize variants of clinical significance (VCSs), variants of uncertain significance (VUSs) and generated a pilot scheme to calculate genetic scores representing the genetic load of rare and common variants in the synaptic transmission pathway (GSSyT, and GSSyTc). The association with the primary outcomes (parent-reported autistic traits, Social Responsiveness Scale) was computed using a mixed linear model. Behavioral and genetic features were combined in a machine learning (ML) model to predict the individual response within the cohort.

Results

In total, 4.4% (n = 9) and 11.3% (n = 23) of the cohort carried VCSs and VUSs, respectively. Compared to non-carriers, individuals with VCS or VUS together tended to have limited improvements of the interventions (β = 9.22; CI (-0.25, 18.70); P = 0.057) and improved significantly less from standard care (β = 9.35; CI (0.70, 18.00); P = 0.036), but not from SSGT (β = -2.50; CI (-13.34, 8.35); P = 0.65). In addition, both GSSyT, and GSSyTc were associated with differential outcomes in standard care and SSGT groups. Our ML model showed the importance of rare variants for outcome prediction.

Conclusions

Autistic individuals with likely pathogenic rare variants identified by exome sequencing might benefit less from the standard care. SSGT could therefore be of heightened importance for this subgroup. Further studies are needed to understand genetic predisposition to intervention outcomes.

**Trial registration**: Social Skills Group Training ("KONTAKT") for Children and Adolescent With High-functioning Autism Spectrum Disorders (KONTAKT-RCT), NCT01854346.


**Keywords**: autism spectrum disorder, social skills group training, exome sequencing, rare variants
Background

The molecular genetic understanding of autism spectrum disorder (ASD) has undergone breakthroughs in research settings over the last decade, and these findings are continuously translated into clinical use [1,2]. Currently, the main implication of clinical genetic testing for an individual with ASD is to potentially identify a molecular diagnosis that can be used as a diagnosis specifier according to the Diagnostic & Statistical Manual of Mental Disorders (DSM-5) [3]. The standard recommended genetic testing for individuals with ASD has been targeted testing for Fragile X syndrome and chromosomal microarray to detect copy number variations (CNVs) [4]. However, a recent consensus statement suggested that exome sequencing detecting rare genetic variants should replace chromosomal microarray as the first-tier genetic testing for neurodevelopmental disorders (NDDs), including ASD [5]. Despite these recommendations, the utilization of genetic testing in ASD is still low [6], and the usefulness of the genetic information beyond a molecularly defined diagnosis is still limited. In addition to rare genetic variants, polygenic risk score (PRS), which aggregates the risk of common variants from genome-wide association studies (GWAS), is associated with ASD and shows an additive effect with rare variants [7,8]. A justified assumption is that genetic information can help to develop more individualized and effective interventions [9]. Nevertheless, few studies are focused on investigating how genetic information or genetically defined subgroups of autistic individuals can be used to guide the choice or development of treatments and interventions for ASD [10].

Genetic findings in ASD have shown important insights into the biology of the condition by suggesting the convergence of gene networks such as synaptic transmission, gene regulation, and chromatin remodeling pathways [11–13]. Machine learning (ML) prediction models are increasingly being used in relation to ASD, and studies have successfully used genetic information to predict additional genes, diagnosis, and variants pathogenicity [14–16].
However, a few studies have used ML methods to predict intervention outcomes in ASD and other mental disorders by integrating genetic and clinical predictors [17,18].

With challenges in social communication and interaction being core features of ASD, the most common interventions offered to autistic individuals target social skills or development [19]. Social skills group training (SSGT) is a behavioral intervention widely applied in ASD to ameliorate social communication difficulties, with varied responses depending on factors such as age and sex [20]. In addition, our group demonstrated that both rare CNVs and common variant PRS could affect individual intervention outcomes in a randomized controlled trial evaluating the SSGT KONTAKT® [21,22]. Seeking to examine further the role of rare variants on individual outcomes of SSGT and standard care, we performed exome sequencing in the participants and leveraged all level genetic data and clinical information from the KONTAKT® trial [20], as summarized in Figure 1. Additionally, we proposed a pilot scheme on how to construct a gene set specific genetic score measuring the contribution of common and rare variants in synaptic transmission genes to gain biological insight for outcome variability. Lastly, we built a machine learning model combining clinical and genetic features to predict intervention responses within the KONTAKT® randomized controlled trial.

Methods

Participants

All participants in this study were a part of the SSGT KONTAKT® randomized controlled trial (clinical trial identifier: NCT01854346) [20]. The inclusion and exclusion criteria, participant characteristics, and the trial outcome have been reported in detail earlier [20]. In brief, all the participants had a confirmed ASD diagnosis, had an IQ > 70, and at least one comorbid neurodevelopmental or psychiatric diagnosis (such as attention deficit hyperactivity
disorder (ADHD), anxiety disorder, or depression). In KONTAKT®, participants are trained in groups of children (7-13 years) or adolescents (13-17 years). In the trial, participants were randomly assigned to either the standard care receiving any usually provided clinical intervention, or the SSGT group receiving standard care plus SSGT. A total of 207 (SSGT group: 105, standard care group: 102) participants were included in this study, who had consented to the genetic part and had available saliva samples as well as recorded parent-reported Social Responsiveness Scale (SRS) [23] total scores (primary outcome) at baseline and either at post-intervention (n = 186) or at 12-week follow-up (n = 169).

The four subscales of the SRS probing different social behavioral traits (social awareness, social cognition, social motivation, social communication) were also used in this study as targeted outcomes [23]. As a part of the analysis presented in this study, we determined a binary classification outcome based on the change in the SRS from pre- to post-treatment. Participants with a >10% decrease in the total SRS score were classified as "response to treatment". Additional information on the other clinical measures can be found in Supplementary Information.

**Variant calling and prioritization**

Exome sequencing was done on NovaSeq 6000 using paired-end 2 × 150 readout. Detailed information on library preparation and quality control can be found in Supplementary Information. After sequencing, data quality control, reads alignment, variant calling, and annotations were completed using Mutation Identification Pipeline (MIP) [24] (https://github.com/Clinical-Genomics/MIP). The average sequencing coverage of OMIM genes in all subjects was 77.9, and 98.2% OMIM gene exons reached 10X sequence depth (Supplementary Fig 1). In addition, 89.9% and 88.5% of the exome were covered with at least 10X and 20X sequence depth, respectively.
The called variants were first filtered based on quality and minor allele frequency (Supplementary Information). The variant effect, defined by the Sequence Ontology (http://www.sequenceontology.org/), was annotated by Ensembl Variant Effect Predictor (VEP) v92. After quality control, qualified rare variants with the effect of loss-of-function (LOF, splice donor/acceptor, frameshift, start lost, stop gained) or damaging missense were prioritized as rare deleterious variants for clinical significance assessment. Rare deleterious variants that were located in developmental disorder-related genes (Supplementary Table 1, see Supplementary Information), and in accordance with the gene inheritance pattern were considered developmental disorder-related rare variants. Furthermore, evaluation to categorize the developmental disorder-related rare variants to pathogenic/likely pathogenic, uncertain clinical effect, or benign/likely benign based on ACMG guideline was done [25]. Both pathogenic and likely pathogenic variants were classified as variants of clinical significance (VCSs). All VCSs and variants of uncertain significance (VUSs) were validated by Sanger sequencing.

**Genetic score for synaptic transmission (GSSyT)**

To test a pilot scheme for analyzing gene set specific genetic load, we constructed genetic scores for synaptic transmission (SyT) genes (GSSyT). To calculate GSSyT based on the rare genetic variants (GSSyTr), we used the number of genes with rare deleterious exome variants in the synaptic transmission gene set and brain gene set combining with rare CNV genes [21] (Supplementary Information). The number of synaptic genes captured by two kinds of rare variants was added and divided by the number of brain-expressed genes with rare variants in each individual. Next, we generated GSSyT for the common variants (GSSyTc) acquired from the previous study [22]. The calculation of GSSyTc was based on set-based polygenic risk score (PRS) using a new feature of PRSice named PRSet.
We chose the largest ASD GWAS of Grove et al. [7] as reference data to calculate set-based PRS for ASD. The ratio between synaptic transmission and brain PRSs was represented as GSSyT_c.

**Statistical analyses**

We applied a mixed linear model to assess whether VCS/VUS carrier status was associated with intervention outcome measured by parent-reported SRS total score and its four subscales. We included a three-way interaction: carrier status*time points*interventions together with all lower-order interactions, age, sex, PRS for ADHD (P-threshold 1.0), clinically significant CNVs as fixed factors, and the clinics in which the individuals had participated in the intervention and individual IDs as random factors in the model. In addition, to measure the separate effect of VCS/VUS on SSGT or standard care, the two-way interactions of carrier status*time points were tested using the same mixed linear model in SSGT and standard care subgroups separately. Furthermore, we also performed linear regression as a secondary model to test whether VCS/VUS and ASD PRS (P-threshold 0.5) independently affect the intervention outcomes.

Similarly, we conducted a mixed linear model to investigate the effect of the variants from the synaptic transmission pathway on the intervention outcomes. We combined two three-way interactions of GSSyT_r*time points*interventions and GSSyT_c*time points*interventions, together with age, sex as fixed factors, and same random factors as earlier. We also tested the effect of GSSyT_c and GSSyT_r in two intervention subgroups using the two-way interaction of GSSyT*time points. Furthermore, linear regression was performed to test the interaction of GSSyT_c and GSSyT_r using the change in SRS total score and sex and age as cofactors.

**Machine learning prediction**
As an outcome for our machine learning prediction, a binary outcome for response to treatment (>10% decrease in the total SRS score) or not was calculated. The prediction model was constructed using the participant characteristics, comorbidity diagnoses, treatments, autism-related scales assessed at pre-treatment, and genetic information of rare and common variants. Detailed feature information is listed in Supplementary Table 2. We performed nested cross-validation by splitting total data into training and validation sets using outer cycle (10-fold) and then splitting the training set again into training and test data using inner cycle (5-fold). A linear support vector machine (SVM) was employed for prediction, and recursive feature elimination was implemented to rank the most important features. We used the area under the receiver operating characteristic curve (AUC) as model performance measurement. Other metrics, such as sensitivity, specificity, and accuracy, were evaluated at each outer cycle cross-validation.

More detailed information can be found in Supplementary Information.

Results

Clinically significant variants and phenotype characteristics

Exome sequencing data were analyzed for 207 individuals from the SSGT KONTAKT® randomized controlled trial. Three individuals were excluded due to the low sample quality. In the remaining data from 204 individuals, 102 (50.0%) were from the SSGT group, 60 (29.4%) were female, and 124 (60.8%) were adolescents. Based on our variant prioritization scheme, each individual had on average 54 rare deleterious variants, of which four were putatively developmental disorder-related rare variants (Supplementary Fig 2). From these, we identified nine variants of clinical significance (VCSs) in nine carriers (4.4%) and 24 variants of uncertain significance (VUSs) in 23 carriers (11.3%) following ACMG guideline [25] (Table 1,
Supplementary Table 3). All VCSs (five indels, four single nucleotide variations (SNVs)) were rare loss-of-function variants in autosomal dominant genes (Table 1). In total, 29 individuals had either one VCS or VUS, and two individuals carried two different variants (one individual with a VUS and a VCS, and another individual with two VUSs) respectively. There were no significant differences in parental reported total Social Responsiveness Scale (SRS) score at pre-treatment (P = 0.68), IQ (P = 0.86), or Autism Diagnostic Observation Schedule (ADOS) social communication total scores (P = 0.28) between individuals with VCS/VUS and those without these variants (Supplementary Fig 3). We also tested the difference after removing individuals with earlier reported clinically significant CNVs [21], and found similar non-significant results (pre-treatment SRS: P = 0.69, IQ: P = 0.82, ADOS: P = 0.30).

In addition to exome sequencing, we used our earlier acquired data on clinically significant copy number variations (CNVs) and common variants polygenic risk score (PRS) for ASD [21,22]. The distribution of PRS for ASD did not differ between individuals with or without VCS/VUS from exome sequencing (P = 0.37, Supplementary Fig 4a). Two individuals were carriers of both clinically significant CNVs and VUSs, but no individual had both clinically significant CNVs and VCSs (P = 0.65, Supplementary Fig 4b). Detailed distribution of PRS for ASD and carriers with different prioritized variants were shown in Supplementary Fig 4c.

Association between carrier status and SRS score

To investigate the role of VCS/VUS identified in this study on the SSGT and standard care intervention outcomes, we used parent-reported autistic traits total score on the SRS as the primary outcome and tested interactions of carrier status of prioritized variants, time points, and interventions using a mixed linear model in the whole cohort. We first analyzed individuals with VCS or VUS together in comparison with non-carriers. The interaction of carrier status
and time points indicated that individuals who were carriers tended to improve less at post-
intervention ($\beta = 9.22, P = 0.057$, Supplementary Table 4, Fig 2). Further analysis of the
changes in the four SRS subscales (social communication, social awareness, social cognition,
social motivation) showed that the total SRS score differences were driven by changes in the
social cognition subscale (($VCS + VUS)_{\text{post-intervention}}: \beta = 3.34, P = 0.006$,
Supplementary Fig 5). The other three subscales did not show any significant associations with
the carrier status (Supplementary Fig 5). A similar trend was observed when comparing carriers
of VCS to non-carriers. We found VCS carriers had significantly less improvements at post-
intervention ($\beta = 14.84, P = 0.029$, Supplementary Table 4), especially for the changes of social
cognition subscale ($VCS_{\text{post-intervention}}: \beta = 4.38, P = 0.0095$). No outcome difference was
detected between VUS carriers and non-carriers ($VUS_{\text{post-intervention}}: \beta = 4.36, P = 0.49$,
Supplementary Table 4). We did not find any significant association at follow-up either
(Supplementary Table 4). When contrasting the two intervention groups in three-way
interaction, no significant association was observed with SRS total scores or any of the four
subscales (Fig 2, Supplementary Table 4, Supplementary Fig 5). Next, to distinguish the
influence of the VCS/VUS within each intervention group, we performed the association
analysis for the intervention groups separately. When combining VCS and VUS carriers into a
single carrier group, significantly worse outcome was found in carriers receiving standard care
compared to non-carriers at post-intervention ($\beta = 9.35, P = 0.036$, Supplementary Table 5).
Similarly, VCS carriers in the standard care group also improved less than non-carriers at post-
intervention ($\beta = 14.86, P = 0.017$, Supplementary Table 5). No significant association was
found in the SSGT group (Supplementary Table 5).

We also tested the interaction of PRS for ASD from our previous study [22] and
VCS/VUS in the linear model. Outcomes at post-intervention and follow-up were measured by
SRS changes at two time points compared to pre-intervention. Similarly, we found carriers of
VCS had less improvements at post-intervention (VCS: $\beta = 14.66$, $P = 0.03$). No significant associations were found either in the two-way (exome variant*PRS) or the three-way interactions (exome variant*PRS*interventions), demonstrating the independent effect of PRS and rare pathogenic exome variants on treatment outcomes (Supplementary Fig 6).

**Synaptic transmission gene set on intervention outcomes**

Based on our earlier findings of synaptic gene sets and the intervention outcome [22], we generated a pilot scheme to calculate gene set specific scores based on both rare (rare deleterious exome variants and rare CNVs) and common variants from synaptic transmission genes and assessed their effect on interventions. Protein coding genes expressed in the brain were used as a background gene set. We calculated genetic scores in synaptic transmission gene set (GSSyT) for each individual to separately represent the genetic load of rare (GSSyT_r) and common variants (GSSyT_c), as shown in Supplementary Fig 7. There was no correlation between the scores generated from the rare and common variants affecting synaptic transmission genes (Pearson r = 0.069, $P = 0.35$) and no statistical difference of GSSyT_r and GSSyT_c between two intervention groups (GSSyT_r: $t = 0.65$, $P = 0.51$; GSSyT_c: $t = 0.11$, $P = 0.91$). The GSSyT_c had no association with the intervention outcomes in the whole cohort (Supplementary Table 6). However, individuals with higher GSSyT_c in the SSGT group had better outcomes compared with the standard care group at post-intervention ($\beta = -5.52$, $P = 0.033$, Fig 3). On the contrary, we found a significant outcome improvement at follow-up in individuals with higher GSSyT_r in the whole cohort ($\beta = -5.38$, $P = 0.016$, Supplementary Table 6). Further, individuals with higher GSSyT_r had less improvements after SSGT compared with standard care at follow-up ($\beta = 8.30$, $P = 0.0044$, Fig 3). After testing GSSyT effect in each intervention group, we found similar associations where higher GSSyT_r was beneficial for outcome within the standard care ($\beta = -5.46$, $P = 0.0091$, Supplementary Table 7), and higher
GSSyT\textsubscript{c} was associated with favorable outcome after SSGT (\(\beta = -5.42, P = 0.0062\), Supplementary Table 7).

To further test the relationship between GSSyT\textsubscript{r} and GSSyT\textsubscript{c}, we also used linear regression adding the interaction effect of GSSyT\textsubscript{r} and GSSyT\textsubscript{c} together with interventions using SRS changes of post-intervention and follow-up. Similar to previous results, improved effects of both GSSyT\textsubscript{r} and GSSyT\textsubscript{c} were indicated on interventions (post-intervention: GSSyT\textsubscript{r}*SSGT: \(\beta = -5.95, P = 0.028\); follow-up: GSSyT\textsubscript{r}: \(\beta = -6.67, P = 0.014\), GSSyT\textsubscript{r}*SSGT: \(\beta = 9.32, P = 0.0077\)). However, we did not detect any significant interaction (Supplementary Fig 8).

**Intervention outcome prediction using machine learning test**

Lastly, we investigated whether a potential machine learning (ML) model could predict the response by integrating detailed clinical and genetic information from the individuals in the KONTAKT\textsuperscript{®} trial. Participants were separated into two classes based on if they had responses to the treatment (\(n = 88\) (47.3%)) or not (\(n = 98\) (52.7%)) measured by parent-reported SRS at post-intervention. After evaluation, we applied linear support vector machine (Supplementary Information) on selected twelve features, including individual comorbidities, SSGT treatment, SRS and Clinical Global Impression (CGI) score, rare exome variants (number of rare damaging/developmental disorder-related variants), rare CNVs (number of genes in rare CNVs, rare CNV size) and GSSyT\textsubscript{r} (Supplementary Fig 9). The average area under the receiver operating characteristic curve (AUC) and accuracy were 0.632 and 0.634 (Fig 4). The ML model was able to correctly predict individuals' responses to interventions with an average of 60% sensitivity and 66.4% specificity (Fig 4).

**Discussion**
Most of the current interventions yield variable responses among autistic individuals. There is minimal knowledge of whether genetic information can be used to guide clinicians to individualize intervention decisions. A few case studies [26–28] and our earlier reports indicate a potential to identify an association between intervention outcome and genetic information [21,22]. Despite large-scale exome sequencing studies in ASD, no studies have directly connected ASD-related rare sequence variants to intervention outcomes. Here, we investigated whether carrier status of rare ASD-related sequence variants, categorized as variants of clinical significance (VCSs), variants of uncertain significance (VUSs), or variants in genes within synaptic transmission pathway, have an effect on outcomes following SSGT and standard care from a recent clinical trial. After variant prioritization and the association test, we found that individuals with VCS or VUS together showed limited social behavioral change across the entire cohort and further analysis pinpointed that this effect was especially evident for the standard care group. Furthermore, we showed that the effect was driven by changes in the social cognition subscale of the SRS. Additionally, the limiting effect of VCS was stronger than VUS in both the entire cohort and the standard care. As SSGT was an add-on intervention to the standard care in the KONTAKT® trial, the non-significant difference between carriers and non-carriers in the SSGT group might indicate a potentially beneficial effect of SSGT for carriers compared to standard care alone, which should be explored further. Earlier, we demonstrated that individuals with clinically significant CNVs had better outcomes in standard care and less improvements after SSGT [21]. Interestingly, in our pilot-scheme of genetic scores based on variant effects on genes within synaptic transmission pathway (GSSyT), we similarly showed that rare variant GSSyT (GSSyT_r) and common variant GSSyT (GSSyT_c) were associated with differential outcomes in standard care and SSGT groups. Further our ML model confirmed the importance of rare variants for outcome prediction and provided a feasible
pattern for investigating the role of both clinical and genetic features on ASD-related treatments.

We also provided evidence that intellectually able autistic individuals carried genetic variants of clinical significance, as we detected that a minority of 4.4% individuals had VCS, which is within the range of other clinical exome sequencing studies [29–31]. Although the proportion is slightly lower than in earlier reports, it is important to note that these studies also included individuals with co-occurring intellectual disability, which is associated with an increased probability of VCS findings [29–31]. Combining clinically significant CNVs with the exome sequencing findings, the overall proportion of identifiable genetic etiology of autistic individuals in this cohort was 12.6%, demonstrating that this group of autistic individuals should also have access to clinical genetic testing. As no overlap was found between VCS and clinically significant CNV carriers, our results confirmed the complementary role of exome sequencing and chromosome microarray increasing the molecular diagnostic yield for individuals with ASD.

We demonstrated nine VCSs from different autosomal-dominant genes implicated in ASD and other NDDs. For example, variants from CUL3 (OMIM* 603136) have been observed in individuals with pseudohypoaldosteronism (MIM* 614496) and contribute to ASD, social deficits, and anxiety-like behaviors [32–34]. Likewise, we found both clinically significant CNV and exome VCS affected ASH1L (OMIM* 607999) in our samples related to intellectual disability (MIM*617796), ASD, and Tourette syndrome [30,35]. This gene encodes a histone methyltransferase that associates with the transcribed region such as Hox genes, catalyzes H3K36 methylation, and plays important roles in neurodevelopment [36].

When assessing the distribution of rare and common variants in this cohort, we found that carriers of pathogenic variants and non-carriers had identical levels of PRS for ASD, which is in line with a recent study using a much larger cohort showing that the contribution from
common genetic variants is similarly relevant to autistic individuals with or without a strongly acting de novo mutation [8]. In addition, Niemi et al. compared PRS between children with severe NDDs (e.g., global developmental delay, intellectual disability, cognitive impairment or learning disabilities, and ASD) with and without pathogenic variants and found no significant differences for any tested PRSs [37]. All these results suggest PRS and rare variants associated with ASD or other NDDs may contribute additively to the individual liability for the conditions. Similarly, we found no interaction effect of PRS for ASD and exome pathogenic variants on interventions. As a similar independent role for PRS and clinically significant CNV was shown earlier [22], our study indicated an additive or differential role for each type of variant contributing to the treatment outcomes.

There is a high genetic and clinical heterogeneity in ASD. Besides traditional association tests, an ML model can integrate measurements to recapitulate useful information to predict intervention outcomes. Using the KONTAKT® trial data, we built a prediction model to select the most contributing features to the intervention responses, including variables representing the complexity and severity of the disorder (comorbidity status, SRS and CGI scores), intervention conditions (SSGT intervention), and genetic liability especially rare variant related characteristics. These selected features from different domains could provide not only a prioritized list for future ASD investigations but more evidence for precise guidance into clinical practice of intervention decisions. When the linear model, especially the linear SVM, combined selected variables into outcomes, it outperformed the non-linear model showing 63.4% accuracy, illustrating the outcome was to a certain degree predictable instead of by chance (50%) to help clinicians decide whether an individual could have a response or not after adding SSGT as an add-on of standard care. Due to limited power, our study was inadequate to make predictions for each intervention subgroup. As no study has combined clinical and genetic factors to predict the outcomes of ASD interventions, the performance of
prediction in our study could be considered a reference to compare after similar studies become available.

Limitations in our study should also be noticed. To begin with, although our study was based on the largest randomized controlled trial of SSGT in ASD to date, limited sample size restricted the power to identify intervention-associated genes and to stratify variant effect in subgroups. Exome sequencing studies of ASD have shown that both de novo and inherited rare variants are major components of genetic risk and are associated with cognitive phenotypes such as IQ [38]. However, due to the lack of parental information, our study could not distinguish de novo and inherited rare variants and seek out any differences between variant types. Furthermore, as the treatment methods are varied for ASD, we could not combine samples from other clinical cohorts and validate our findings. Although some completed and ongoing clinical trials testing the SSGT effect, exome sequencing data on the same cohorts are unavailable at present [39,40]. Therefore, without further validations, our results must be interpreted with caution. With increased genetic testing in clinics, converting genetic information into primary diagnosis and treatment, combined with clinical observation and biometric data, could proactively shape clinical decisions and long-term management [41]. Factors from different aspects should be considered when leveraging genetic data into practice, such as data interpretation, genetic counseling, ethical consideration, and economic constraints [41,42].

Conclusion

In summary, we have demonstrated that exome sequencing can identify molecular diagnoses in an ASD cohort without intellectually disabled individuals, and genetic information reveals significant associations with intervention outcomes, although with variable effects in the SSGT and standard care groups. We also show that gene set based genetic scores could be further
explored to determine meaningful subgroups and direct future studies on biological mechanisms. By developing an ML model, we confirm that individual outcomes could be predicted by clinical and genetic factors such as rare variants. Our study is a preliminary exploration of rare variants and a combination of different level variants on ASD intervention outcomes. Future clinical trials of different interventions in ASD should include genetic data collection to improve the use of molecular genetics beyond diagnoses for individualized intervention plans.

List of abbreviations

ASD: autism spectrum disorder
CNV: copy number variation
NDD: neurodevelopmental disorder
PRS: polygenic risk score
ML: machine learning
SSGT: social skills group training
ADHD: attention deficit hyperactivity disorder
SRS: Social Responsiveness Scale
VCS: variant of clinical significance
VUS: variant of uncertain significance
GSSyT: genetic score for synaptic transmission
GSSyTc: genetic score for synaptic transmission for common variant
GSSyr: genetic score for synaptic transmission for rare variant
SVM: support vector machine
AUC: area under the receiver operating characteristic curve
Declarations

Ethics approval
Written informed consent was obtained from parents or legal guardians, and verbal consent from all participants. All the protocols and methods were followed in accordance with the Declaration of Helsinki and were approved by the ethical review board in Stockholm (Dnr 2012/385-31/4) and the clinical authorities of the two involved counties.

Availability of data and materials
All public data from databases and studies were described in the manuscript. The genetic data is accessible through the Swedish National Data Service (https://snd.gu.se/en) after necessary clearances. Code related to variants prioritization and developmental disorder related variants selection is available at: https://github.com/DanyangLi107/vcf_varselect. Code related to machine learning test is available at: https://github.com/DanyangLi107/ml_kontakt. All statistical analyses were performed in R v3.6.3. Additional software and package information were illustrated in the Methods and Supplementary Information.

Competing interests
Sven Bölte is an author of the German and Swedish KONTAKT manuals and receives royalties from Hogrefe publishers. Dr. Bölte discloses that he has in the last 3 years acted as an author, consultant or lecturer for Medice and Roche. Nora Choque Olsson and Ulf Jonsson are authors of the Swedish KONTAKT manuals. Martin Becker is a full-time employee of Bayer AG, Germany. The other authors do not report financial interests or potential conflicts of interest.

Funding
This work was supported by grants from the Swedish Research Council clinical therapy framework grant (921-2014-6999, Drs. Bölte, Tammimies), the Swedish Research Council, in partnership with the Swedish Research Council for Health, Working Life and Welfare, Formas and VINNOVA (cross-disciplinary research program concerning children's and young people's mental health, 259-2012-24, Dr. Bölte), Stockholm County Council (20130314 Dr. Bölte, 20170415 Dr. Tammimies), Swedish Foundation for Strategic Research (ICA14-0028, FFL18-0104, Dr. Tammimies), The Swedish Brain Foundation (Dr. Tammimies), the Harald and Greta Jeansson Foundations (Dr. Tammimies), Åke Wiberg Foundation (Dr. Tammimies), StratNeuro (Dr. Tammimies), the L'Oréal-UNESCO for Women in Science prize in Sweden with support from the Young Academy of Sweden (Dr. Tammimies), Sällskapet Barnavård (Dr. Tammimies, Ms Li), China Scholarship Council (Ms Li), Drottning Silvias Jubileumsfond (Ms Li) and Committee of Research at Karolinska Institutet (Dr Tammimies).

Authors' contributions

D.L and K.T conceived and planned the present study. D.L designed and performed the analyses with supervision from K.T. H.J and N.N provided overall suggestion and discussion for the analyses. M.B and A.A performed experiments of prioritized variants for Sanger validation. S.B and N.C-O conceived the original clinical trial. S.B is responsible for the clinical data and provided the expertise of the clinical measures together with N.C-O, and U.J. D.L and K.T wrote the paper with input from all authors. All authors approved the final version of the article.

Acknowledgments

We thank the children, adolescents, and parents who participated in the study. Christina Coco, Oskar Flygare, Anders Görling, and Kerstin Andersson are acknowledged for their work in
collecting the data and samples during the RCT. The authors are also thankful for the leads of
child and adolescent psychiatry (Olav Bengtsson, Paula Liljeberg, Charlotta Wiberg
Spangenberg, Peter Ericson, Karin Forler, Alkisti Nikolayidis Linderholm, all of Stockholm
County Council), PRIMA Järva child and adolescent psychiatry (MaiBritt Giacobini), and
child and adolescent habilitation services (Lars Kjellin, Moa Pellrud, of Örebro County Council)
for organizational support. The DNA extraction was done at the KI Biobank. We thank Anna
Gellerbring, Emma Sernstad, Emilia Ottosson and Valtteri Wirta from Clinical Genomics
Stockholm core facility, Karolinska Institutet, and Science for Life Laboratory for exome
sequencing experiments and variant calling process. The computation resources were provided
by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science
(UPPMAX). We would like to acknowledge the Swedish Bioinformatics Advisory Program at
the National Bioinformatics Infrastructure Sweden, SciLifeLab, for bioinformatics
consultation for data analysis.

Reference
1. Finucane BM, Ledbetter DH, Vorstman JA. Diagnostic genetic testing for
neurodevelopmental psychiatric disorders: closing the gap between recommendation and
contributions to autism spectrum disorder. Psychol Med . Cambridge University Press; 2021;1–
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders
Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals
analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical


http://doi.wiley.com/10.1002/ajmg.a.38193


Figure legends

Figure 1. Overview of the study and performed analysis. SSGT: social skills group training; ADHD: attention deficit hyperactivity disorder; ADOS: Autism Diagnostic Observation Schedule; CNV: copy number variation; SNP: single nucleotide polymorphism; GSSyT: genetic score for synaptic transmission genes.

Figure 2. Association between VCS+VUS carrier status and intervention outcomes. a. Least-squares means of Social Responsiveness Scale (SRS) among carriers at pre-, post-intervention and follow-up. b. Least-square means of SRS among carriers at pre-, post-intervention and follow-up in standard care and social skills group training (SSGT) groups. VCS: variant of clinical significance; VUS: variant of uncertain significance.

Figure 3. The effect of rare and common variant genetic scores for synaptic transmission genes (GSSyT_r and GSSyT_c) on intervention outcomes at post-intervention and follow-up. The color of each circle represents the coefficient value in 3-way interaction of GSSyT_r/GSSyT_c*time points*interventions. The size of each circle is the coefficient -log10p value calculated from the mixed linear model.

Figure 4. Performance of machine learning model measured by area under the receiver operating characteristic curve (AUC), accuracy, sensitivity and specificity.
Table 1. Detailed information of the prioritized clinically significant variants identified in the cohort.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Chr</th>
<th>Position</th>
<th>Reference allele</th>
<th>Alternative allele</th>
<th>Genotype</th>
<th>Effect</th>
<th>Transcript ID</th>
<th>Exon</th>
<th>CDS position change</th>
<th>蛋白 change</th>
<th>Transcripts affected/Total transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASH1L</td>
<td>1</td>
<td>155307491</td>
<td>C</td>
<td>CT</td>
<td>Het</td>
<td>FRAMESHIFT</td>
<td>ENST00000392403</td>
<td>28/28</td>
<td>c.8854-8855insT</td>
<td>p.R2952X</td>
<td>2/3</td>
</tr>
<tr>
<td>CHD1</td>
<td>5</td>
<td>98262076</td>
<td>ACT</td>
<td>A</td>
<td>Het</td>
<td>FRAMESHIFT</td>
<td>ENST00000284049</td>
<td>1/35</td>
<td>c.13-14del</td>
<td>p.S5X</td>
<td>1/1</td>
</tr>
<tr>
<td>MAPK8IP3</td>
<td>16</td>
<td>1808160</td>
<td>GGTGA</td>
<td>G</td>
<td>Het</td>
<td>SPLICE DONOR</td>
<td>ENST00000250894</td>
<td>9/31</td>
<td>-</td>
<td>-</td>
<td>2/2</td>
</tr>
<tr>
<td>KANSL1</td>
<td>17</td>
<td>44249260</td>
<td>AGAGAT</td>
<td>A</td>
<td>Het</td>
<td>FRAMESHIFT</td>
<td>ENST00000262419</td>
<td>2/15</td>
<td>c.245-249insGAGAT</td>
<td>p.YL82-83X</td>
<td>7/9</td>
</tr>
<tr>
<td>PACS1</td>
<td>11</td>
<td>65838072</td>
<td>C</td>
<td>T</td>
<td>Het</td>
<td>STOP GAINED</td>
<td>ENST00000320580</td>
<td>1/24</td>
<td>c.C115T</td>
<td>p.Q39*</td>
<td>1/8</td>
</tr>
<tr>
<td>ZBTB20</td>
<td>3</td>
<td>114058002</td>
<td>AG</td>
<td>A</td>
<td>Het</td>
<td>FRAMESHIFT</td>
<td>ENST00000462705</td>
<td>10/10</td>
<td>c.1856del</td>
<td>p.P619X</td>
<td>7/9</td>
</tr>
</tbody>
</table>

All coordinates were based on human reference genome build GRCh37/ hg19. The transcripts were based on Ensembl GRCh37.

Abbreviations:
- Chr: chromosome
- CDS: coding sequence
- Het: heterogeneous

CC-BY 4.0 International license
This preprint has been licensed under a Creative Commons Attribution 4.0 International License. It is made available under a Creative Commons Attribution 4.0 International License (which was not certified by peer review). The copyright holder for this preprint is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

doi: 10.1101/2021.05.19.21257395
medRxiv preprint
Figure 1. Overview of the study and performed analysis. SSGT: social skills group training; ADHD: attention deficit hyperactivity disorder; ADOS: Autism Diagnostic Observation Schedule; CNV: copy number variation; SNP: single nucleotide polymorphism; GSSyT: genetic score for synaptic transmission genes.
Figure 2. Association between VCS+VUS carrier status and intervention outcomes. a. Least-squares means of Social Responsiveness Scale (SRS) among carriers at pre-, post-intervention and follow-up. b. Least-square means of SRS among carriers at pre-, post-intervention and follow-up in standard care and social skills group training (SSGT) groups. VCS: variant of clinical significance; VUS: variant of uncertain significance.
Figure 3. The effect of rare and common variant genetic scores for synaptic transmission genes (GSSyT_r and GSSyT_c) on intervention outcomes at post-intervention and follow-up. The color of each circle represents the coefficient value in 3-way interaction of GSSyT_r/GSSyT_c*time points*interventions. The size of each circle is the coefficient -log10p value calculated from the mixed linear model.
Figure 4. Performance of machine learning model measured by area under the receiver operating characteristic curve (AUC), accuracy, sensitivity and specificity.