

The influence of common polygenic risk and gene sets on social skills group training response in autism spectrum disorder

Danyang Li^{1,2}, Nora Choque-Olsson PhD^{1,3}, Hong Jiao PhD⁴, Nina Norgren PhD⁵, Ulf Jonsson PhD^{1,6}, Sven Bölte PhD^{1,7,*,#}, Kristiina Tammimies PhD^{1,2,*,#}

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

³Center for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet and Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

⁴Department of Biosciences and Nutrition, Karolinska Institutet, and Clinical Research Centre, Karolinska University Hospital, Huddinge, Sweden

⁵Department of Molecular Biology, National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Umeå University, 901 87 Umeå, Sweden

⁶Department of Neuroscience, Child and Adolescent Psychiatry, Uppsala University, Sweden

⁷Curtin Autism Research Group, School of Occupational Therapy, Social Work and Speech Pathology, Curtin University, Perth, Western Australia

*equal contribution

#corresponding author

Information of the corresponding author:

Kristiina Tammimies

Full postal address: Bioclinicum J9:30, Visiongatan 4, 171 56 Solna, Sweden

Telephone: +46762379741

E-mail address: kristiina.tammimies@ki.se

Sven Bölte

Full postal address: KIND/CAP Research Center, Gävlegatan 22B, 8th floor, 11330 Stockholm, Sweden

Telephone: +46851452220

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

Abstract

Social skills group training (SSGT) is one of the most frequently used behavior interventions in children and adolescents with autism spectrum disorder (ASD). Current evidence suggests that the effects are moderate and heterogeneous. Genetic predisposition could be one of the factors contributing to this heterogeneity. Therefore, we used polygenic risk score (PRS) and gene-set analysis to investigate the association between SSGT response and common variants in autistic individuals. Participants from the largest randomized clinical trial of SSGT in ASD to date were selected for genotyping. Polygenic risk scores (PRSs) for ASD, attention deficit hyperactivity disorder (ADHD), and educational attainment (EA) were calculated, and their associations with the intervention outcome at post-intervention and follow-up were tested using mixed linear model. In addition, thirty-two gene sets within five categories (synaptic, glial, FMRP, glutamate, and mitochondrial) were selected to evaluate their role in the intervention outcome. Individuals with higher PRSs for ASD and ADHD had inferior response after SSGT. After multiple test correction, significant results were kept for higher ADHD PRS at follow-up ($\beta = 6.67$, $p = 0.016$). Five gene sets within synaptic category showed modest association with reduced response to SSGT in ASD. Taken together, we provided preliminary evidence that genetic liability calculated using PRS and common variants in synapse gene sets could influence the outcome of SSGT. Our results hold promise for future research into the genetic contribution to individual response to ASD interventions, and should be validated in larger cohorts.

Introduction

The response to behavioral interventions varies between individuals with the same neurodevelopmental and psychiatric disorders. One of the factors contributing to this heterogeneity might be genetic predisposition. Genome-wide association studies (GWAS) have been used to pinpoint common variants, with mostly small effects, associated with disorders or personality traits across a genome. For behavioral interventions, the size of available cohorts with detailed outcomes is limited, and to date no significant variants have been identified to be associated with outcome for any specific interventions using GWAS. However, the use of polygenic risk score (PRS), an aggregate measure of the cumulative effects of single nucleotide polymorphisms (SNPs) derived from GWAS, has provided promising results in psychiatry both for behavioral and pharmacological treatments¹⁻⁵. For instance, PRS has been studied to predict the response to cognitive behavior therapy in major depressive disorder². In addition to PRS, gene-set analysis can be utilized to group multiple genetic variants in genes and further to related gene sets to unravel biological process and cellular functions related to intervention responses. Relevant gene set associations have been identified for interventions in psychiatry⁶⁻¹⁰. For example, genetic variations in genes underlying glutamatergic or NMDA neurotransmission have been implicated in short-term antipsychotic medication efficacy in schizophrenia and calcium signaling pathway has been indicated in response to selective serotonin reuptake inhibitors in obsessive-compulsive disorder^{8,10}.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication and interaction, together with restricted, repetitive behaviors¹¹. ASD commonly co-exists with other neurodevelopmental and psychiatric disorders such as attention-deficit hyperactivity disorder (ADHD), anxiety, depression, and intellectual disability¹². The genetic knowledge in ASD has increased rapidly in recent years. For numbers of the individuals with ASD, rare genetic variants, such as loss-of-function variants and copy number variations (CNVs) in specific genes and loci, can indicate a molecular etiology¹³. In addition, common variants have been shown to be of causal significance in the disorder^{14–16}. Although there are only few genome-wide significant SNPs identified for ASD, the cumulative polygenic variation summarized by PRS has shown to be predictive of ASD and autistic traits in the general population^{17,18}. Studies have also indicated shared genetic liability of polygenic risk in ASD and psychiatric disorders^{19,20}. Educational attainment (EA), defined as the highest degree of education, also has a confirmed genetic correlation with ASD²¹ and academic achievement has been linked to social skills²². Both rare and common genetic variations in ASD have shown to converge to specific gene-sets such as synaptic formation and targets of the Fragile-X mental retardation protein (FMRP)^{23,24}. To date, there are no studies investigating how the already implicated common genetic variation in ASD would relate to intervention outcomes in autistic individuals.

Social skills group training (SSGT) is one of the most frequently used behavioral interventions in ASD, aiming to alleviate social communication difficulties in verbal

individuals within normative intellectual range in a group setting. The largest randomized controlled trial (RCT) of SSGT (“KONTAKT”) to date, conducted by our center in Sweden, included children and adolescents with ASD and at least one common neurodevelopmental or psychiatric comorbidity²⁵, such as ADHD, anxiety and depression. SSGT as an add-on to standard care was found to have a small to moderate effect compared to standard care only, with significant effects on the primary outcome limited to adolescents (13-17 years) and females²⁵.

We recently showed that autistic individuals who were carriers of clinically significant and rare genic CNVs larger than 500kb had significantly inferior outcomes after SSGT within the RCT²⁶. Here, we expand our genetic investigations to test the association between SSGT intervention response in ASD and common variants using PRS and gene-set analysis. We hypothesized that PRSs for ASD, ADHD, and EA, as well as common variants within known ASD gene sets, would influence the response to SSGT. Additionally, we tested if there was a significant correlation with the clinical measures of SRS, ADHD diagnosis and IQ in our clinical ASD cohort with the selected PRS. To the best of our knowledge, this is the first study to evaluate the influence of common variants on SSGT response in autism or any other neurodevelopmental disorders. Our results will provide further evidence of potential use of genetic profiles to predict individual-level outcomes for ASD interventions.

Methods

Study individuals

The original multicenter, randomized pragmatic RCT of SSGT (“KONTAKT”) recruited participants from 13 child and adolescent psychiatry outpatient units in Sweden between August 2012 and October 2015. A detailed description of the inclusion and exclusion criteria have been earlier described by Choque-Olsson et al²⁵. In short, 296 children (7–12 years) and adolescents (13–17 years) with a diagnosis of autism, atypical autism, Asperger syndrome, or pervasive developmental disorder not otherwise specified using ICD-10 criteria were included in the trial²⁷. All participants had full-scale IQ > 70 according to the Wechsler Intelligence Scale for Children and at least one common comorbid psychiatric diagnosis of ADHD, depression, or anxiety disorder according to ICD-10^{27,28}. During the 12-weeks trial, the standard care group (n=146) received any ongoing support or intervention, and the remaining 150 participants were included in the active SSGT condition. The parent-reported Social Responsiveness Scale (SRS) assessing autistic-like traits was used as the primary intervention outcome measure. The SRS is a 65-item Likert-type scale generating totals scores ranging between 0 and 195, with higher score indicating greater autism trait severity²⁹. SRS was recorded for each individual at baseline (pre-intervention), 12 weeks (post-intervention), and 3-month after the end of the intervention (follow-up).

Participants who contributed saliva samples and had the primary outcome measure recorded at either post-intervention or follow-up were included in this study. After selection, clinical data and samples from 207 participants (SSGT group: 105, standard care group: 102) were used for genotyping as described in elsewhere²⁶.

Ethics

Written informed consent from the parents or legal guardians and verbal consent from the children and adolescents was collected. All the protocols and methods in this study were in accordance with the Declaration of Helsinki. The trial and sample collection were approved by the ethical review board in Stockholm (Dnr 2012/385-31/4) and the clinical authorities of the two involved counties. The trial was registered online (NCT01854346).

Genotyping and quality control

DNA collection, extraction, and genotyping procedures are described elsewhere in detail²⁶. In short, the genotyping was done on the Affymetrix CytoScan™ HD microarray platform, containing 743 304 SNP probes, in two separate batches. Data from genotyping were transformed from Affymetrix .CEL format to .tped format using “Affy2sv” package v1.0.14 in R. Position of each SNP was located based on microarray annotation reference file (version NA32.3). Quality control (QC) of the data using PLINK v1.90 was performed on per-individual basis within each genotyping batch to remove poorly genotyped individuals and on per-marker QC to exclude low-quality markers following a published protocol^{30,31}. Individuals with discordant sex, heterozygosity rate > 3SD, individual genotype failure rate > 0.03, and relatedness were removed. Ancestry of the participants was estimated using principle component analysis (PCA) in EIGENSOFT v7.2.1. We restricted our analyses to European ancestry.

Thus, individuals were excluded based on two largest principle components (principle component 1 < 0.00 , principle component 2 < 0.05). Additionally, the largest four PCAs were added in the statistical model to adjust for ancestry. As no batch effects were detected, the qualified data were combined to clean low-quality markers with following criteria: minor allele frequency < 0.05 , Hardy-Weinberg equilibrium $< 1e-06$, individual missingness < 0.1 , and marker missingness < 0.05 .

Imputation

We used the 1000 Genomes phase III haplotype data including all ancestries as a reference genome. SNPs passing QC were separated into autosomes, and haplotypes were inferred based on reference panel using SHAPEIT v2³². For each phased autosome, imputation was performed in 5Mb windows using IMPUTE2 v2.3.2^{33,34}. All imputed regions were combined for post-imputation QC. Imputed SNPs were filtered using following metrics: info score < 0.8 , minor allele frequency < 0.05 , Hardy-Weinberg equilibrium $< 1e-06$, marker missing rate < 0.05 , and individual missing rate < 0.1 . SNPs were intersected together after post-imputation QC using both SNPTEST v2.5.5 and PLINK v1.90^{31,35}.

Polygenic Risk Score calculation

To calculate the PRSs for ASD, ADHD, and EA, we used the latest GWAS summary results from Psychiatric Genomics Consortium^{16,36,37}. To minimize different population effects, we included only individuals with European ancestry from the GWAS reference

samples. The estimated odds ratio and P-value of each SNP allele were used from each reference set. SNPs in both reference and our in-house data were pruned using clumping with a cutoff of $r^2 \geq 0.1$ within 500 kb window. PRS was calculated based on independent SNPs using five P-value thresholds (Pt) (< 0.01 , < 0.05 , < 0.1 , < 0.5 , < 1) selected from three reference sets, with higher Pt incorporating more SNP effects. Allele numbers at each SNP (0, 1, 2), weighted by the natural logarithm of the allelic odds ratio were summed to calculate an accumulative effect across the genome^{1,38}. PRS was then standardized (mean = 0, SD = 1) to test for association. All calculations were conducted using PRSice v2.1.4³⁹.

Gene sets generation

Reference gene sets were obtained based on a previously published study²⁴. Thirty-two gene sets within five categories: synaptic (1047 genes), glial (240 genes), FMRP (1809 genes), glutamate (156 genes), and mitochondrial (132 genes) were included. SNPs were annotated to genes based on European population from 1000 genomes and gene locations, Build 37.

Statistical analysis

PRS association

We used mixed linear model (MLM) to identify the effect of PRS on the SSGT response. Three time-points (pre-, post-intervention, and follow-up), as well as two study groups (SSGT and standard care) were included to test the interaction effect of

PRS*time*intervention. Based on our previous studies^{25,26}, factors associated with inferior intervention outcomes were younger age (children), male gender, and CNV carrier status (large size CNVs (> 500kb) and clinically significant CNVs). Therefore, age group, gender, and population stratification of the four largest PCAs were added as fixed factors in the model. Additionally, the clinical centers and each individual were used as random factors. Carrier status of large size CNVs and clinically significant CNVs were also added as fixed factors in separate models comparing with the model without CNV adjustment. Marginal R^2 for fixed effect and conditional R^2 for fixed and random effects were assessed to explain the role of PRS on the variance of treatment outcome. For each phenotype on each Pt, both marginal and conditional R^2 were derived from the difference between model with PRS and model without PRS, performed by “MuMIn” package v1.43.6 in R⁴⁰. Beta and 95% confidential interval (CI) were estimated from MLM to evaluate the interaction effect of PRS with SSGT and standard care at different times. After false-discovery rate multiple corrections among three PRSs, original P-values under 0.0167 were considered significant⁴¹. In addition, two-sided student’s t-test and Pearson correlation were used to examine the correlation between PRS and closely related characteristics, including PRS for ADHD and ADHD comorbidity status, PRS for EA and IQ level, and PRS for ASD and ASD severity (SRS score at pre-treatment). All analyses were performed in R v3.4.2.

Gene sets association

MAGMA v1.0.6 was used to perform gene and gene-set analysis⁴². Linear regression

was chosen to identify the effects of common variation within specific gene sets on intervention outcome. The changes of SRS score between post-intervention or follow-up and pre-intervention were used as regression outcomes. Age, gender, four largest PCAs and CNV carrier status (large size CNVs and clinically significant CNVs) were added in the model as cofactors. The estimated effect size of competitive test and P-value on each gene set were obtained followed by multiple testing correction with positive effect size indicating a decreased effect of the SSGT response.

Results

Sample characteristics

Samples of 188 participants passed genotyping QC, of which 99 belonged to active SSGT. When comparing the individuals of this RCT subgroup with the total cohort from the RCT²⁵, we did not detect any differences in characteristics between these individuals and those not passing genotyping QC ([Supplementary Table 1](#)). The final genotype data sets consist of 539 106 SNPs after genotyping marker QC and 5 126 694 SNPs after imputation QC.

Correlation between PRS and related characteristics

The correlation between ADHD comorbidity status and PRS for ADHD, pre-intervention SRS and PRS for ASD, as well as IQ and PRS for EA were tested ([Supplementary Figure 1](#)). No correlation was observed in ASD PRS and baseline SRS ([Supplementary Figure 1A](#)). However, autistic participants with ADHD comorbidity

had higher ADHD PRS compared with participants without ADHD (Pt = 0.1, $p = 0.010$; Pt = 0.5, $p = 0.0075$; Pt = 1.0, $p = 0.0077$) (Supplementary Figure 1B), and IQ was positively correlated with PRS for EA (Pt = 0.05, $p = 0.038$; Pt = 0.1, $p = 0.024$; Pt = 0.5, $p = 0.014$; Pt = 1.0, $p = 0.012$) (Supplementary Figure 1C).

Proportion of variance in SSGT outcome explained by PRS

For the model examining the effect of PRS for ASD, the largest explained marginal and conditional variance in the SSGT outcome was Pt 0.5 when controlling for the clinically significant CNVs (Figure 1). We detected only small differences in the explained variance when testing for ADHD PRS at all Pt (Figure 1). For EA PRS, increasing values of marginal and conditional R^2 occurred with higher Pt, in which Pt 1.0 explained the most variance. All three tested models showed similar results (Figure 1, Supplementary Figure 2).

Association between PRS and SSGT outcome

The PRS calculated from the Pt which had the highest explained variance for each phenotype is shown as main result (Figure 2). Associations for all Pts in three models are shown in Supplementary Table 2. Inferior outcomes after SSGT were implicated in individuals with higher PRS for ASD on Pt 0.5 ($\beta = 6.47$, $p = 0.019$) and Pt 1.0 ($\beta = 5.93$, $p = 0.031$) at follow-up. However, the associations were not significant after multiple testing correction. For ADHD PRS, participants with higher score also improved less from SSGT on Pt 0.05 (post-intervention: $\beta = 6.22$, $p = 0.019$, follow-up:

$\beta = 6.93$, $p = 0.012$), Pt 0.1 (post-intervention: $\beta = 6.22$, $p = 0.017$, follow-up: $\beta = 5.75$, $p = 0.036$) at both post-intervention and follow-up and Pt 0.5 ($\beta = 6.00$, $p = 0.030$), Pt 1.0 ($\beta = 6.67$, $p = 0.016$) at follow-up. After multiple testing correction, significant association remained in the follow-up for ADHD PRS score at Pt 0.05 and Pt 1.0. No significant effects were found for EA PRS.

Gene sets association with SSGT response

We tested if common genetic variation in five gene set groups and 31 ASD related gene sets could explain SSGT outcome using competitive gene-set analysis. For the outcome at post-intervention, four gene sets showed modest effect to cause significant inferior outcome: intracellular signal transduction (large size CNVs: $\beta = 0.204$, $p = 0.0027$; clinically significant CNVs: $\beta = 0.202$, $p = 0.0029$), cell adhesion and trans-synaptic signaling (large size CNVs: $\beta = 0.255$, $p = 0.0071$; clinically significant CNVs: $\beta = 0.247$, $p = 0.0090$), excitability (large size CNVs: $\beta = 0.268$, $p = 0.017$; clinically significant CNVs: $\beta = 0.247$, $p = 0.026$), GPCR signaling (large size CNVs: $\beta = 0.255$, $p = 0.023$; clinically significant CNVs: $\beta = 0.252$, $p = 0.024$), all belonging to synaptic group (Table 1). In follow-up, only RNA and protein synthesis, folding and breakdown (RPSFB) (large size CNVs: $\beta = 0.184$, $p = 0.030$; clinically significant CNVs: $\beta = 0.185$, $p = 0.030$) in synaptic group showed similar effect (Table 1). However, none of the gene set results remained significant after multiple testing correction. The results for other gene sets did not reach nominal significance (Supplementary Table 3).

Discussion

In this study, by calculating PRS and analyzing related gene sets, we tested if different subsets of common variants and biological gene groups would influence the outcome of SSGT. We showed that a higher common genetic variant load for ADHD and ASD was negatively associated with the SSGT response in autistic individuals, albeit with modest effects compared to rare CNVs impact we have shown before²⁶. However, the effect of PRS remained even when we controlled for rare CNV carrier status of the individuals suggesting an independent role of PRS on the intervention outcome. We also demonstrated significant correlations between PRS for ADHD and ADHD comorbidity and PRS for EA and IQ level in our study cohort. Additionally, we showed suggestive evidence that genes involved in synaptic functions could be important for modulating the response to SSGT, as several gene sets important for synaptic functions showed nominal significance to influence the SSGT outcome.

Given the high heterogeneity of responses to SSGT among individuals with ASD, it is important to find factors influencing the outcome, which could be used to tailor the interventions for each individual. Our results provided intriguing evidence that in addition to the rare high impact CNVs, the polygenic contribution of common variants could have a moderating role on SSGT outcome, especially the ADHD PRS. Highly shared genetic heritability between ASD and ADHD has been reported in family studies^{43,44}, and genetic overlap between the two neurodevelopmental conditions is evident from genome-wide correlation analysis^{21,45,46}. A recent study has identified that

PRS for ADHD can be used to predict ASD-related measures such as pragmatic language abilities⁴⁷. Compared to ASD alone, having ASD and ADHD symptoms were confirmed to be associated with greater impairments in socialization adaptive skills in clinical presentation⁴⁸. Based on genetic sharing between the two conditions, our results highlight the connection between ADHD genetic information and ASD intervention effects. Furthermore, the correlation between PRS and ADHD comorbidity indicates that ADHD PRS could be used to evaluate liability to ADHD in individuals first diagnosed with ASD.

Compared to the results of PRS for ADHD, we found only indicative association between PRS for ASD and intervention outcomes. Currently, it is not known if the same genetic factors would influence the risk of developing ASD and intervention response measured here. In other disorders such as major depressive disorder (MDD), studies have found variants associated with MDD can be either positive or negative for the outcome of antidepressants⁴⁹, and the results of PRS as a predictor of treatment outcome are inconsistent^{50,51}. Further studies should clarify how ASD PRS might modulate intervention outcomes.

We did not find any association between PRS for EA and SSGT intervention response. Studies using genetic correlation analysis corroborated interrelated results between ASD and years of education and college attainment^{21,52}. Some PRS studies suggest that PRS for ASD is associated with EA⁵², and EA PRS is associated with lower

externalizing behaviors⁵³, but no study has shown the prediction of autistic traits or ASD-related interventions using PRS for EA. As expected, we found a positive correlation between PRS EA and IQ in our sample, which is in accordance with other findings in the general population⁵⁴.

Interestingly, the inferior outcome for individuals with higher PRS for ASD and ADHD seemed to be particularly pronounced at follow-up. A similar tendency was previously also seen in individuals with clinically significant rare CNVs²⁶. This suggest that additional care during or after SSGT treatment should be considered for individuals with higher genetic risk, as to alleviate the genetic influence on the intervention outcome. Recently, a long version of SSGT intervention (24-week) was conducted resulting in larger positive effects compared to the 12-week intervention⁵⁵. Further studies should investigate if genetic effects remain when the intervention dose is increased.

Although no robust association after multiple testing corrections were shown for the gene-set analysis, we found suggestive evidence that genes necessary for synapse formation and maintenance may influence SSGT response. Since synapse formation and synaptic plasticity have been indicated as one of the key neuronal mechanisms in ASD, differences in these key steps in brain development could possibly play a major role in how individuals respond to interventions⁵⁶. Some aspects of treatment-induced behavior improvement are also related to brain plasticity changes in psychiatric

disorders. For instance, significant increases of grey matter in left hippocampus and left amygdala correlated with the degree of improved cognition are found in early-onset schizophrenia individuals after two years of social skills group therapy and cognitive remediation⁵⁷.

We acknowledge the limitations of our study. First, the sample size from RCT is limited, which restricted the precision of our analyses and precluded in-depth analysis of complex pathways and single variants. Secondly, the analyzed sample included individuals with normative IQ and common neurodevelopmental and psychiatric comorbidities, limiting the generalizability across the total population of autistic individuals. Finally, studies have shown loci harboring common alleles are also enriched for rare variants with large effects from whole-exome sequencing in psychiatric disorders⁵⁸. Although we have adjusted for the carrier status of rare CNVs, other rare variants were not controlled for. In the future, more genetic information can be combined to better understand their effect on SSGT outcome.

Conclusion

This is the first study showing that common polygenic risk, especially for ADHD, is associated with SSGT response in ASD individuals, and gene sets from synaptic groups may play a potential biological role in SSGT response. Replications including larger sample size and combination of more genetic and clinical factors are needed to further clarify the genetic influences and mechanisms behind individual intervention responses.

Web resources

1000 Genomes Phase III haplotypes reference data:

https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference

Psychiatric Genomics Consortium: <https://www.med.unc.edu/pgc/>

Gene locations from the NCBI site and SNP locations from 1000 genomes Phase 3

European population Build 37 for MAGMA: <https://ctg.cncr.nl/software/magma>

Acknowledgments

We thank the children, adolescents, and parents who participated in the study. Christina Coco MSc, Oskar Flygare, MSc, Anders Görling MSc, and Kerstin Andersson MSc are acknowledged for their work in collecting the data and samples during the RCT. We also thank Ielyzaveta Rabkina and Sofia Stamouli for the help with handling the DNA samples and genetic data. The authors are also thankful to the leads of child and adolescent psychiatry (Olav Bengtsson, MD, Paula Liljeberg, MD, Charlotta Wiberg Spangenberg, MSc, Peter Ericson, MSc, Karin Forler, MSc, Alkisti Nikolayidis Linderholm, MSc, all of Stockholm County Council), PRIMA Järva child and adolescent psychiatry (MaiBritt Giacobini, MD, PhD), and child and adolescent habilitation services (Lars Kjellin, PhD, Moa Pellrud, MSc, of Örebro County Council) for organizational support. We would like to acknowledge Martin Becker and Ielyzaveta Rabkina for their contributions to the genetic work. The DNA extraction was done at the KI Biobank. The genotyping was done at Aros Applied Biotechnology in

Denmark. The computation resources were provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). We also want to thank the Swedish Bioinformatic Advisory program for providing bioinformatic consultation for data analysis.

Contributions

D.L and K.T conceived and planned the presented study. D.L performed the analysis with supervision from N.N, and K.T. H.J. provided overall suggestion and discussion for the analyses. S.B and N.C-O conceived the original clinical trial. S.B is responsible for the clinical data and provided 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 paper.

Conflicting of interest

Sven Bölte is an author of the German and Swedish KONTAKT manuals and receives royalties from Huber/Hogrefe publishers. Dr. Bölte discloses that he has in the last 5 years acted as an author, consultant or lecturer for Shire/Takeda, Medice, Roche, Eli Lilly, Prima Psychiatry, and SB Education and Psychological Consulting AB. He receives royalties for text books and diagnostic tools from Huber/Hogrefe, Kohlhammer and UTB. Nora Choque-Olsson and Ulf Jonsson are authors of the Swedish KONTAKT manuals. 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, Dr. Tammimies), The Swedish Brain Foundation (Dr. Tammimies), the Harald and Greta Jeansson's 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 Board of Research at Karolinska Institutet (Dr Tammimies).

References

- 1 Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 2009; **18**: 3525–3531.
- 2 Andersson E, Crowley JJ, Lindefors N, Ljótsson B, Hedman-Lagerlöf E, Boberg J *et al.* Genetics of response to cognitive behavior therapy in adults

- with major depression: a preliminary report. *Mol Psychiatry* 2019; **24**: 484–490.
- 3 Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Heilbronner U *et al.* Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: A genome-wide association study. *JAMA Psychiatry* 2018; **75**: 65–74.
- 4 Zhang JP, Robinson D, Yu J, Gallego J, Fleischhacker WW, Kahn RS *et al.* Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. *Am J Psychiatry* 2019; **176**: 21–28.
- 5 Alemany-Navarro M, Costas J, Real E, Segalàs C, Bertolín S, Domènech L *et al.* Do polygenic risk and stressful life events predict pharmacological treatment response in obsessive compulsive disorder? A gene–environment interaction approach. *Transl Psychiatry* 2019; **9**. doi:10.1038/s41398-019-0410-0.
- 6 de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. *Nat Rev Genet* 2016; **17**: 353–364.
- 7 Wigmore EM, Hafferty JD, Hall LS, Howard DM, Clarke T-K, Fabbri C *et al.* Genome-wide association study of antidepressant treatment resistance in a population-based cohort using health service prescription data and meta-analysis with GENDEP. *Pharmacogenomics J* 2019; : 1.
- 8 Wang Q, Man Wu H, Yue W, Yan H, Zhang Y, Tan L *et al.* Effect of Damaging Rare Mutations in Synapse-Related Gene Sets on Response to

- Short-term Antipsychotic Medication in Chinese Patients With Schizophrenia.
JAMA Psychiatry 2018; **75**: 1261.
- 9 De Jong S, Vidler LR, Mokrab Y, Collier DA, Breen G. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. *J Psychopharmacol* 2016; **30**: 826–830.
- 10 Umehara H, Numata S, Tajima A, Nishi A, Nakataki M, Imoto I *et al*. Calcium Signaling Pathway Is Associated with the Long-Term Clinical Response to Selective Serotonin Reuptake Inhibitors (SSRI) and SSRI with Antipsychotics in Patients with Obsessive-Compulsive Disorder. *PLoS One* 2016; **11**: e0157232.
- 11 Association AP. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5), Fifth Edition*. American Psychiatric Publishing: Arlington, VA, 2013.
- 12 Lai M-C, Lombardo M V, Baron-Cohen S. Autism. *Lancet* 2014; **383**: 896–910.
- 13 Tamimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC *et al*. Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *JAMA* 2015; **314**: 895.
- 14 Iakoucheva LM, Muotri AR, Sebat J. Getting to the Cores of Autism. *Cell* 2019; **178**: 1287–1298.
- 15 Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RJL, Nurnberger Jr JI, Hallmayer JF. Autism genetics: opportunities and challenges for clinical

- translation. *Nat Rev Genet* 2017; **18**: 362–376.
- 16 Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H *et al.*
Identification of common genetic risk variants for autism spectrum disorder.
Nat Genet 2019; **51**: 431–444.
- 17 Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H *et al.*
Identification of common genetic risk variants for autism spectrum disorder.
Nat Genet 2019; **51**: 431–444.
- 18 Bralten J, Van Hulzen KJ, Martens MB, Galesloot TE, Arias Vasquez A,
Kiemeny LA *et al.* Autism spectrum disorders and autistic traits share genetics
and biology. *Mol Psychiatry* 2018; **23**: 1205–1212.
- 19 Guo W, Samuels JF, Wang Y, Cao H, Ritter M, Nestadt PS *et al.* Polygenic
risk score and heritability estimates reveals a genetic relationship between ASD
and OCD. *Eur Neuropsychopharmacol* 2017; **27**: 657–666.
- 20 Pourcain BS, Robinson EB, Anttila V, Sullivan BB, Maller J, Golding J *et al.*
ASD and schizophrenia show distinct developmental profiles in common
genetic overlap with population-based social communication difficulties. *Mol
Psychiatry* 2018; **23**: 263–270.
- 21 Brainstorm Consortium TB, Anttila V, Bulik-Sullivan B, Finucane HK,
Walters RK, Bras J *et al.* Analysis of shared heritability in common disorders
of the brain. *Science* 2018; **360**: eaap8757.
- 22 Gustavsen AM. Longitudinal relationship between social skills and academic
achievement in a gender perspective. *Cogent Educ* 2017; **4**.

- doi:10.1080/2331186X.2017.1411035.
- 23 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- 24 Jansen A, Dieleman GC, Smit AB, Verhage M, Verhulst FC, Polderman TJC *et al.* Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *Eur J Hum Genet* 2017; **25**: 863–868.
- 25 Choque Olsson N, Flygare O, Coco C, Görling A, Råde A, Chen Q *et al.* Social Skills Training for Children and Adolescents With Autism Spectrum Disorder: A Randomized Controlled Trial. *J Am Acad Child Adolesc Psychiatry* 2017; **56**: 585–592.
- 26 Tammimies K, Li D, Rabkina I, Stamouli S, Becker M, Nicolaou V *et al.* Association between Copy Number Variation and Response to Social Skills Training in Autism Spectrum Disorder. *Sci Rep* 2019; **9**: 9810.
- 27 Organization WH. *International Classification of Diseases and Related Health Problems 10th Revision*. World Health Organization: Geneva, 1992.
- 28 Wechsler D. *The Wechsler intelligence Scale for Children—Fourth Edition*. Pearson Assessment: London, 2004.
- 29 Constantino JN, Gruber CP. *Social Responsiveness Scale (SRS)*. Western Psychological Services: Los Angeles, 2005.
- 30 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*

- 2010; **5**: 1564–1573.
- 31 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al*.
PLINK: A Tool Set for Whole-Genome Association and Population-Based
Linkage Analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 32 Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for
disease and population genetic studies. *Nat Methods* 2013; **10**: 5–6.
- 33 Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype
Imputation Method for the Next Generation of Genome-Wide Association
Studies. *PLoS Genet* 2009; **5**: e1000529.
- 34 Van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg M V., Slagboom PE,
De Bakker PIW *et al*. Population-specific genotype imputations using minimac
or IMPUTE2. *Nat Protoc* 2015; **10**: 1285–1296.
- 35 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint
method for genome-wide association studies by imputation of genotypes. *Nat
Genet* 2007; **39**: 906–913.
- 36 Martin J, Walters RK, Demontis D, Mattheisen M, Lee SH, Robinson E *et al*.
A Genetic Investigation of Sex Bias in the Prevalence of Attention-
Deficit/Hyperactivity Disorder. *Biol Psychiatry* 2018; **83**: 1044–1053.
- 37 Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M *et al*. Gene
discovery and polygenic prediction from a genome-wide association study of
educational attainment in 1.1 million individuals. *Nat Genet* 2018; **50**: 1112–
1121.

- 38 Consortium TIS. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- 39 Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015; **31**: 1466–1468.
- 40 Bartoń K. MuMIn: Multi-Model Inference. 2019. <https://cran.r-project.org/package=MuMIn>.
- 41 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 1995; **57**: 289–300.
- 42 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput Biol* 2015; **11**: e1004219.
- 43 Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The Genetics of Autism Spectrum Disorders and Related Neuropsychiatric Disorders in Childhood. *Am J Psychiatry* 2010; **167**: 1357–1363.
- 44 Ronald A, Simonoff E, Kuntsi J, Asherson P, Plomin R. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. *J Child Psychol Psychiatry* 2008; **49**: 535–542.
- 45 Zhao H, Nyholt DR. Gene-based analyses reveal novel genetic overlap and allelic heterogeneity across five major psychiatric disorders. *Hum Genet* 2017; **136**: 263–274.
- 46 Manley G. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.

- 47 Martin J, Hamshere ML, Stergiakouli E, O'Donovan MC, Thapar A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol Psychiatry* 2014; **76**: 664–71.
- 48 Avni E, Ben-Itzhak E, Zachor DA. The Presence of Comorbid ADHD and Anxiety Symptoms in Autism Spectrum Disorder: Clinical Presentation and Predictors. *Front Psychiatry* 2018; **9**. doi:10.3389/fpsyt.2018.00717.
- 49 Mitjans M, Serretti A, Fabbri C, Gastó C, Catalán R, Fañanás L *et al.* Screening genetic variability at the CNR1 gene in both major depression etiology and clinical response to citalopram treatment. *Psychopharmacology (Berl)* 2013; **227**: 509–519.
- 50 García-González J, Tansey KE, Hauser J, Henigsberg N, Maier W, Mors O *et al.* Pharmacogenetics of antidepressant response: A polygenic approach. *Prog Neuro-Psychopharmacology Biol Psychiatry* 2017; **75**: 128–134.
- 51 Uher R, Tansey KE, Henigsberg N, Wolfgang M, Mors O, Hauser J *et al.* Common Genetic Variation and Antidepressant Efficacy in Major Depressive Disorder: A Meta-Analysis of Three Genome-Wide Pharmacogenetic Studies. *Am J Psychiatry* 2013; **170**: 207–217.
- 52 Hagenaaars SP, Harris SE, Davies G, Hill WD, Liewald DCM, Ritchie SJ *et al.* Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Mol Psychiatry* 2016; **21**: 1624–1632.

- 53 Jansen PR, Polderman TJC, Bolhuis K, van der Ende J, Jaddoe VWV, Verhulst FC *et al.* Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *J Child Psychol Psychiatry Allied Discip* 2018; **59**: 39–47.
- 54 Plomin R, von Stumm S. The new genetics of intelligence. *Nat Rev Genet* 2018; **19**: 148–159.
- 55 Jonsson U, Olsson NC, Coco C, Görling A, Flygare O, Råde A *et al.* Long-term social skills group training for children and adolescents with autism spectrum disorder: a randomized controlled trial. *Eur Child Adolesc Psychiatry* 2019; **28**: 189–201.
- 56 Gilbert J, Man HY. Fundamental elements in autism: From neurogenesis and neurite growth to synaptic plasticity. *Front. Cell. Neurosci.* 2017; **11**.
doi:10.3389/fncel.2017.00359.
- 57 Cramer SC, Sur M, Dobkin BH, O’Brien C, Sanger TD, Trojanowski JQ *et al.* Harnessing neuroplasticity for clinical applications. *Brain.* 2011; **134**: 1591–1609.
- 58 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.

Table 1. Effect of most significant gene sets ($P < 0.05$) on social skill group training (SSGT) at post-intervention and follow-up stratified by clinically significant rare copy number variations (CNVs) and large size ($> 500\text{kb}$) rare CNVs

Gene sets	Gene set groups	Number of genes	Clinically significant CNVs				Large size CNVs			
			Beta	Se	P	Corrected P	Beta	Se	P	Corrected P
Post-intervention										
Cell adhesion and trans-synaptic signaling	Synaptic	74	0.247	0.104	0.0090	0.2340	0.255	0.104	0.0071	0.1988
Excitability	Synaptic	55	0.247	0.127	0.0262	0.5383	0.268	0.127	0.0174	0.4110
GPCR signaling	Synaptic	40	0.252	0.128	0.0244	0.5142	0.255	0.128	0.0231	0.5011
Intracellular signal transduction	Synaptic	138	0.202	0.073	0.0029	0.0837	0.204	0.073	0.0027	0.0832
Follow-up										
RPSFB	Synaptic	62	0.185	0.098	0.0303	0.5871	0.184	0.098	0.0305	0.5928

Abbreviations:

GPCR: G-protein-coupled receptor; RPSFB: RNA and protein synthesis, folding and breakdown

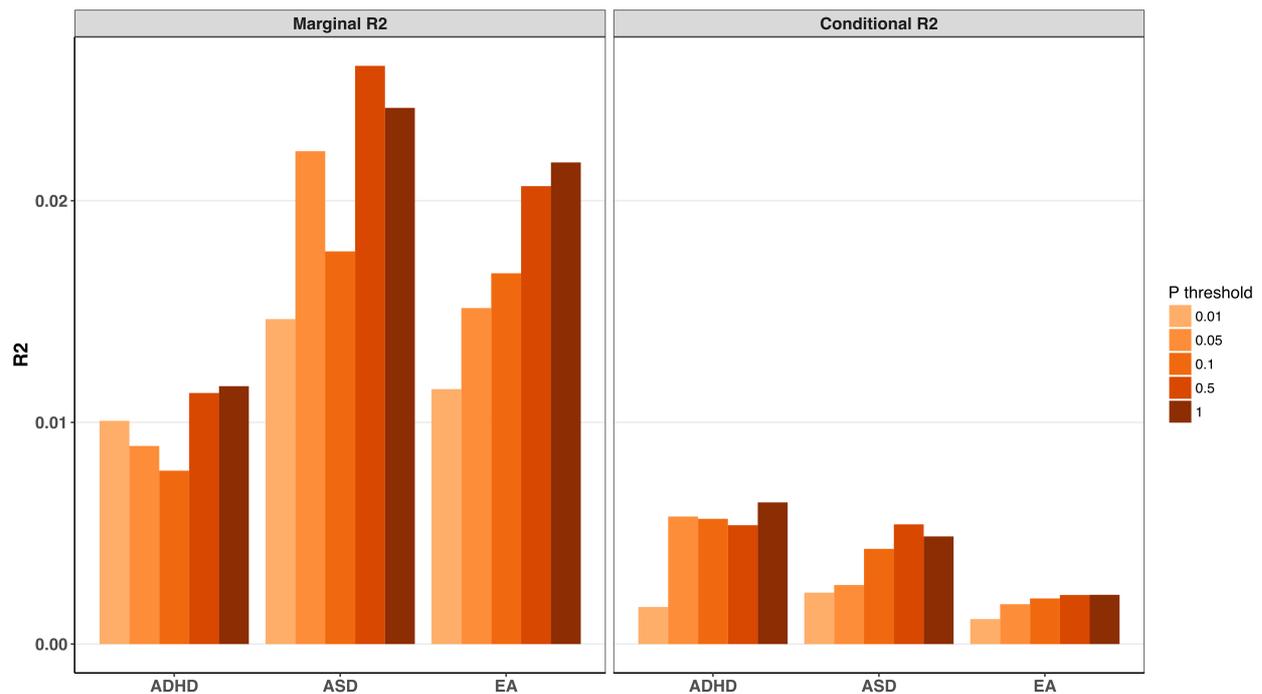


Figure 1. Proportion of variance explained (R^2) by polygenic risk scores (PRSs) of autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and education attainment (EA) in social skill group training (SSGT) derived using five P-value thresholds (< 0.01 , < 0.05 , < 0.1 , < 0.5 , < 1) in the model adjusted for clinical significant rare copy number variation (CNV). Marginal R^2 and conditional R^2 were calculated representing the variance explained by only fixed effects as well as the sum of fixed and random effects.

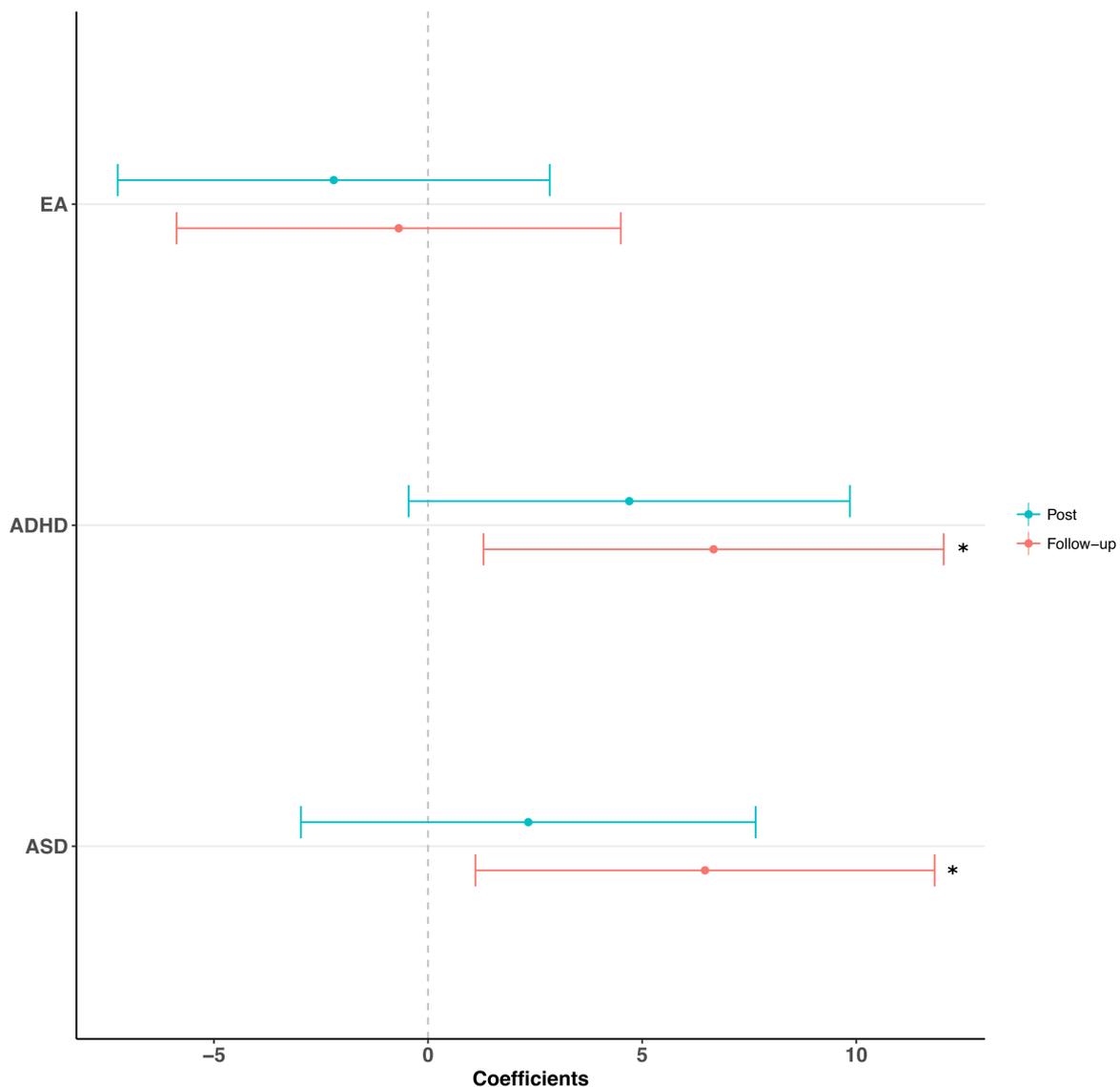


Figure 2. Association of polygenic risk score (PRS) for autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and educational attainment (EA) at post-intervention and follow-up in social skill group training (SSGT). Correlation coefficients are shown with 95% confidence intervals. Different P-value thresholds which have highest explained variance for each PRS were included in the model (ASD: Pt = 0.5, ADHD: Pt = 1, EA: Pt = 1). Large size rare copy number variations (CNVs) (>500kb) and clinically significant rare CNVs were added as cofactors respectively in the model. *: P < 0.05.