The association between a genetic risk score and autistic-like traits.

*Arvid Harder, Psychology stud, Karolinska Institutet*

*Supervisor: Yi Lu, Department of Medical epidemiology and Biostatistics.*

*Nobelsväg 12A*

Abstract

Autism-spectrum disorder (ASD) is a pervasive neurodevelopmental disorder characterized by deficits in social communication, language development and lastly a lack of flexibility in thought patterns and behaviours. It affects roughly 1% of the population. Recent genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) associated with ASD. However, these studies used a case-control model of the disorder, i.e., testing the association among individuals affected and unaffected with ASD. A dimensional approach (i.e., viewing such condition as a normal distribution of the relevant traits, with the tails corresponding to diagnosis has been fruitful in other psychiatric illnesses).

One consequence of ASD reflecting an underlying normal distribution would be that the genes associated with clinical diagnosis of ASD would also be associated with levels of subclinical autistic-like traits in the general population. To investigate this, summary statistics from the latest GWAS was used to generate genetic risk scores (GRS), an individual measure of genetic liability from common variation. We tested if the GRS was associated with subclinical levels of autistic-like traits.

The results demonstrate that the GRS of ASD was significantly associated with autistic-like traits in a young Swedish twin sample. Among the three subdomains, ASD genetic risk scores was most strongly associated with flexibility - questions measuring repetitive and stereotypical behaviour, suggesting that rigid thoughts and behaviour were at the core of the clinical presentation of ASD. Overall, these results indicate that there are shared genetic underpinnings between clinical ASD diagnosis and subclinical levels of autistic-like traits in the general population.

**Introduction**

In an effort to quantify the burden of ASD, Baxter et al estimated that ASD causes a loss of 7.6 million disability-adjusted life years, double that of ADHD and conduct disorder combined (1). The core symptoms of ASD, often called the autism triad are issues with social reciprocity, repetitive and stereotypical behaviour or interests, and impairments in verbal and nonverbal communication. These behaviour symptoms surface before three years of age, and can often be disabling (2). The behavioural genetics revolution has highlighted the roles of genes in the pathogenesis of autism. Twin studies of autism yielded high heritability estimates, ranging between 73-91% (3). The genetic architecture underlying ASD has been studied. It has been shown that half of its genetic liability was explained by common genetic variants (i.e. genetic variants that are common in the population). In 5-15% of cases, the aetiology can be linked to chromosomal rearrangements or single gene disorders (4)(5). Rare mutations were previously thought to play a major role, but recent research has shown that while they may cause significant effects in individuals, collectively rare mutations only explain a small percent of the heritability (4). The liability threshold model hypothesizes that disorders simply reflect the extreme end of an underlying liability which follows a normal distribution, and there is support for this view in autism (6). However, ASD is highly heterogeneous, and its clinical picture varies greatly. Recently, a large GWAS meta-analysis, successfully identified twelve genomic loci associated with ASD, and provided support for differing genetic architectures between subgroups of ASD, e.g., individuals with and without intellectual disability (11). This project sought to extend the published GWAS by understanding how the genetic heterogeneity of ASD is manifested at the symptomatic level. First, we attempted to replicate the known genome-wide significant loci for clinical ASD in a population-based Swedish twin sample, among whom autistic-like traits are measured.

However, looking only at the significant SNPs fails to capture a lot of signal due to the very strict significance levels applied in a GWAS. GRS is a way of capturing more of the information by not only looking at those SNPs that reach significance. Using summary statistics from a GWAS, one can capture the risk each individual carry from alleles common in the population. Using a large twin sample with genotyped data and trait measurements with a set of questions corresponding to each of the autism triad traits, we then derived such a genetic risk score to examine the relationship between risk SNPs and trait measurements of the autism triad.

**Materials and methods**

**CATSS**

The data in this study is part of the Children and Adolescent Twin Study in Sweden (CATSS). Families with twins born in Sweden are invited to join CATSS when twins turn nine or twelve. The overall response rate is 80%. CATSS started in 2004 and is currently ongoing. Phenotypic data were collected via telephone interviews, where one of the parents (in 87.5% percent of cases the mother) answers questions regarding the twins’ mental health and mental development (7).

**Genotype Data**

DNA were extracted from saliva samples at the time of study enrolment. A total of 11551 individuals were available for genotyping using the Illumina PsychChip, out of which 11081 passed quality control. Untyped MZ twin were imputed using genotype data from the paired twin, resulting in a total of 13412 individuals with genotype data. For information on quality control procedure and exclusion criteria, see (11) and (12).

**The Autism-Tics, AD/HD, and other Comorbidities (A-TAC)**

A-TAC is a comprehensive and easy to use screening device for neurodevelopmental disorders. The A-TAC ASD questionnaire has been validated and the reliability has been found to be satisfactory in a sample of clinically diagnosed children, with the area under the ROC curve of 0.95(8). The questionnaire contains 17 questions regarding ASD in three subdomains. The three subdomains include six questions measuring language development, six questions measuring rigidity in actions and thoughts, and five questions measuring social difficulties (8)(9). In total, 12928 individuals with both genotype and phenotype data were available for analyses.

**Genetic Risk Score**

Conceptually, GRS (aka, polygenic risk score) is a weighted sum of all risk variants across the whole genome. It requires a discovery sample of large GWAS, where the effect size and association p-value for each genetic variant were obtained. Here we used the latest GWAS of ASD as *the discovery set* (11). We then calculated the GRS in *the target set* of CATSS samples, by summing up all the risk variants (at a pre-specified p-value) weighted by the effect sizes from the discovery set. The target CATSS samples are independent of the discovery ASD GWAS, i.e. no sample overlap between the two sets.

One important parameter is the p-value cut-off value which is used to select the risk alleles for the GRS calculation. One can use a very stringent cut-off value, such that only the SNPs at genome-wide significance are used to generate the GRS. However, modelling has shown that by using less stringent cut-off values one can capture more signal and obtain a more predictive score. We generated eight sets of GRS with p-values cut-offs of 0.001, 0.01, 0.2, 0.05, 0.1, 0.3, 0.5 and 1. We examined the correlations of these GRS to investigate to what degree more signal is captured. Based on this, we present the main effects on a cut-off level of p < 0.2, and GRS under other p-value cut-offs were tested in a sensitivity analysis (13). Like GWAS, trait association with GRS can be confounded by geographical location - a phenomenon often referred to as population stratification. The most common strategy to control for this is to estimate the principal components (PCs) and adjust those in the association tests (14). Plink, version. 1.9, was used for data processing, quality control, generating PCs, and finally calculating GRS in the CATSS sample.

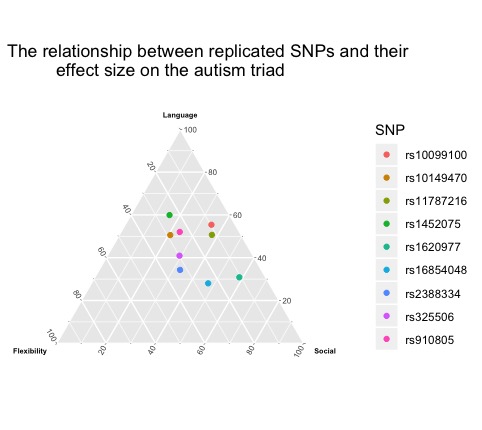
**Analysis**

The analysis was done in R (version 3.6). A linear model was fitted to calculate the association between GRS and autistic-like traits. The effect size can be interpreted as the increase in units of standard deviation on the total scores or subdomains per standard deviation increase in GRS. The estimates on the total score and each subdomain was standardized by dividing each estimate by the standard deviation on the corresponding scale to allow for comparison between estimates. To control for population stratification, we estimated the top twenty PCs. A baseline model consisting of age and all twenty PCs was fitted to predict scores on autistic-like traits. This was done to determine how many PCs should be included in the main analysis. Based on the results of this base model, the first five principal components were selected to be included in the final analysis. The correlations between the scores on each subdomain were also generated. Due to correlated nature of the scores on the subdomains, we applied the false-discovery rate (FDR) to account for multiple comparisons, using the p.adjust function in R. A total of 32 regressions were fitted to determine the relationship between autistic-like traits and the GRS. For each subdomain and the total score, each of the eight GRS were tested yielding 4 x 8 regressions. To control for twin relatedness affecting variance, a sandwich estimator with clustering on each twin pair was used to generate robust standard errors. To calculate variance explained (2) by GRS, we took the difference in2 between the full model with GRS and the baseline model without GRS.

**Results**

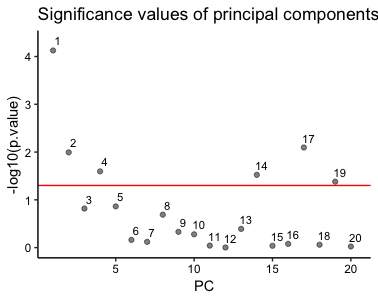
Nine of the twelve loci passed quality control in our sample. We found no obvious pattern of the SNP associations to any of the specific phenotypes of the autism triad (Figure 1).

In the baseline model, the first three were most significant in their association with the total score of autistic-like trait (Figure 2). As a result, we decided to be conservative and adjust for the top five PCs in the final model. The PC14, 17 and 19 were not adjusted because of the limited variance they account for.

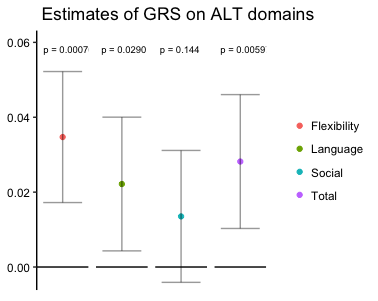
Despite the unclear pattern of individual SNP association, GRS were significantly associated with the total score of autistic-like trait (β [SE] = 0.028 [0.006], FDR-corrected p=0.006) and the two subdomains of flexibility (β [SE] = 0.035 [0.009], FDR-corrected p=0.0008) and language (β [SE] = 0.022 [0.009], FDR-corrected p=0.0290), but not the social domain (FDR-corrected p>0.05) (Figure 3). In the sensitivity analyses, none of the eight GRS at different p-value cut-offs reached significance in its association with the social domain (p> 0.05)(Figure 4). For the total scores, the GRS explained ~ 0.09% of the variance, and for the flexibility domain, the GRS explained ~ 0.13% of the variance.****

**Figure 1:** A ternary plot of the proportion of each effect size in relation to the sum

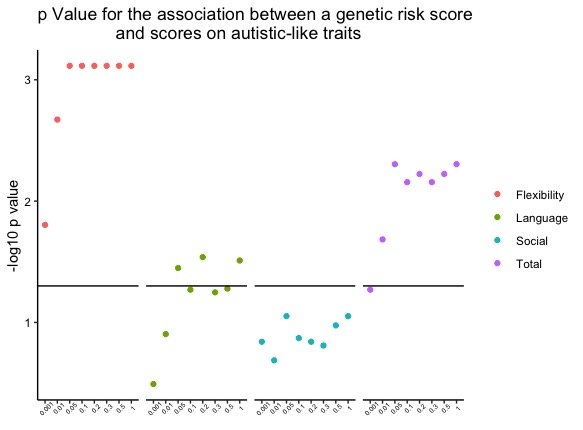
of the three effect sizes.



**Figure 2:** The figure plots the p values of the twenty principal component in a linear model. The p values have been transformed to a negative log10 value. The red line marks p = 0.05.



**Figure 3:** The p-value of the association between risk score and A-TAC measures of the autism triad transformed into –log10.



**Figure 4:** The FDR-adjusted p-values for the association between GRS and autistic-like trait at different p cutoffs. The blackline marks FDR-adjusted significance.

**Discussion**

In this report, wknown forWe first examined the association of the known risk variants for ASD with autistic-like traits and its three subdomains in a Swedish twin population. These known risk variants failed to load to any specific phenotype in the autism triad, which could be due to the limited power in individual variants given their small effect sizes. Therefore, we used a more powerful measure – GRS, which sums all risk variants across the genome weighted by their effects - to test its association with autistic-like trait and its subdomains. We found that the GRS of clinical ASD predicts the levels of autistic-like traits in the general population, albeit only explaining a tiny amount of phenotypic variance. These results support the view that subclinical levels of autistic-like traits share genetic variation with clinical cases of ASD. It is worth noting that the strength of association differs between the subdomains. The domain flexibility was the most significantly related to the GRS. This could indicate that out of the triad of symptoms present in ASD, common genetic variation plays a comparatively larger role in rigid thought patterns and inflexible and stereotypical behaviour than in language issues and social issues. One potential limitation could be that individuals with ASD are the sole drivers of the association, and that the true association between subclinical levels of ASD and the GRS is minimal. However, this has been taken into account in another study, where results were barely impacted by removing clinical cases (10). We therefore find it unlikely that this limitation has driven the observed association.

Acknowledgments

I am very thankful to my supervisor Dr Yi Lu for excellent guidance and mentorship in the process of writing this report. I would also like to thank The Department of Microbiology, Tumors and Cell Biology for organizing KI Summer School in Medical Science, and for making all this possible.

**Citations**

1. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. Psychological Medicine. Cambridge University Press; 2015;45(3):601–13.

2. <https://www.apa.org/topics/autism/>

3. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta‐analysis of twin studies. Journal of Child Psychology and Psychiatry. 2016;57(5):585–95.

4. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nature Genetics. 2014;46(8):881–5.

5. Devlin, Scherer. Genetic architecture in autism spectrum disorder. Current Opinion in Genetics & Development. 2012;22(3):229–37.

6. Robert Plomin, Claire M. A. Haworth, Oliver S. P. Davis. Common disorders are quantitative traits. Nature Reviews Genetics. 2009;10(12):872–8.

7. Anckarsäter H, Lundström S, Kollberg L, Kerekes N, Palm C, Carlström E, et al. The Child and Adolescent Twin Study in Sweden (CATSS). Twin Research and Human Genetics. 2011;14(6):495–508.

8. Hansson, Sara Lina, Svanström Röjvall, Annika, Rastam, Maria, Gillberg, Carina, Gillberg, Christopher, Anckarsäter, Henrik. Psychiatric telephone interview with parents for screening of childhood autism - tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC): preliminary reliability and validity. The British journal of psychiatry : the journal of mental science. 2005;187(3):262–7.

9. Larson, Tomas, Anckarsäter, Henrik, Gillberg, Carina, Ståhlberg, Ola, Carlström, Eva, Kadesjö, Björn, et al. The Autism - Tics, AD/HD and other Comorbidities inventory (A-TAC): further validation of a telephone interview for epidemiological research. BMC Psychiatry. 2010;10(1).

10. Taylor MJ, Martin J, Lu Y, Brikell I, Lundström S, Larsson H, et al. Association of Genetic Risk Factors for Psychiatric Disorders and Traits of These Disorders in a Swedish Population Twin Sample. JAMA Psychiatry. 2019;76(3):280–9.

11. Grove, Jakob, Ripke, Stephan, Als, Thomas D, Mattheisen, Manuel, Walters, Raymond K, Won, Hyejung, et al. Identification of common genetic risk variants for autism spectrum disorder. Nature genetics. 2019;51(3):431–44.

12. Brikell I, Larsson H, Lu Y, Pettersson E, Chen Q, Kuja-Halkola R, et al. The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. Molecular psychiatry. 2018;1–13.

13 Shing Wan Choi, Timothy Shin Heng Mak, Paul F. O’Reilly.

A guide to performing Polygenic Risk Score analyses,

bioRxiv 416545;doi:https//doi.org/10.1101/416545

14. Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, David Reich. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006;38(8):904–9.