CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/121741/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Connolly, Siobhan, Anney, Richard , Gallagher, Louise and Heron, Elizabeth 2019. Evidence of assortative mating in autism spectrum disorder. Biological Psychiatry 86 (4) , pp. 286-293. 10.1016/j.biopsych.2019.04.014

Publishers page: http://dx.doi.org/10.1016/j.biopsych.2019.04.014

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Evidence of Assortative Mating in Autism Spectrum Disorder.

Siobhan Connolly[a1] Richard Anney[ab] Louise Gallagher[a] Elizabeth A. Heron[a]

a : Neuropsychiatric Genetics Research Group,

Dept. of Psychiatry, Trinity College Dublin,

Trinity Centre for Health Sciences, James's

St., Dublin 8, Ireland

b : Institute of Psychological Medicine and Clinical Neurosciences

Cardiff University School of Medicine,

Hadyn Ellis Building, Maindy Road, Cathays,

Cardiff CF24 4HQ, United Kingdom

Number of words in abstract: 239
Number of words in text: 3969
Number of tables: 0
Number of figures: 5
Number of supplementary material: 1
Short title: Evidence of Assortative Mating in ASD

[1]Corresponding author. E-mail: connols4@tcd.ie
Tel: +353(1)8968461

Abstract

Background

Assortative mating is a non-random mating system in which individuals with similar genotypes and/or phenotypes mate with one another more frequently than would be expected in a random mating system. Assortative mating has been hypothesised to play a role in Autism Spectrum Disorder (ASD) in an attempt to explain some of the increase in the prevalence of ASD that has recently been observed. ASD is considered to be a heritable neurodevelopmental disorder but there is limited understanding of its causes. Assortative mating can be explored through both phenotypic and genotypic data, but up until now, has never been investigated through genotypic measures in ASD.

Methods

We investigated genotypically similar mating pairs using genome-wide Single Nucleotide Polymorphism (SNP) data on trio families (Autism Genome Project (AGP) data and Simons Simplex Collection (SSC) data). To determine whether or not an excess in genetic similarity was present we employed kinship coefficients and examined spousal correlation between the principal components in both the AGP and SSC datasets. We also examined assortative mating using phenotype data on the parents to detect any correlation between ASD traits.

Results

We found significant evidence of genetic similarity between the parents of ASD offspring using both methods in the AGP dataset. In the SSC, there was also significant evidence of genetic similarity between the parents when explored through spousal correlation.

Conclusions

This gives further support to the hypothesis that positive assortative mating plays a role in ASD.

Key words: Autism Spectrum Disorder (ASD), assortative mating, genetic assortative mating, ancestral assortative mating, random mating, kinship.

Introduction

Assortative mating occurs when similar males and females mate with each other more (or less) often than expected by chance [1]. This similarity can be trait specific and/or genotype specific [2]. When the trait has a genetic component (phenotype) and assortative mating occurs, then both genetic and phenotypic assortative mating takes place. Ancestral assortative mating occurs when spouse pairs are more (or less) likely to share genes of common ancestry [3]. Studies have largely focused on phenotypic assortative mating, with fewer studies investigating genetic assortative mating [3–8]. Here, we are interested in investigating genetic and phenotypic assortative mating in the complex disorder: Autism Spectrum Disorder (ASD).

ASD is considered to be a heritable [9] neurodevelopmental disorder, characterised by patterns of repetitive behaviours and deficits in language and social behaviour, but there is limited understanding of its causes. Assortative mating has been hypothesised in ASD as parents can often display characteristics of ASD. Baron-Cohen [10] noted estimates of prevalence in 2006 (0.2% [11] and 0.44% [12]) were much higher than the traditional estimate of 0.04% and proposed the hypothesis that this change in prevalence could be due to individuals with ASD-like traits (high-systemisers) mating. A more recent estimate of the prevalence of ASD was much higher with 1 in 68 children aged 8 years having an ASD diagnosis (1.47%) [13]. Peyrot et al. [14] found that, in general, assortative mating could lead to a

considerable increase in prevalence for disorders that are less common and are highly heritability.

Assortative mating has previously been investigated in ASD using a variety of approaches. One such approach has been to look at the educational and occupational phenotypes of parents and family members to see if logical and systematic professions lead to an increase of ASD in the families or an increase in severity of ASD traits among the members of the family [15–19]. A recent study showed that systemising traits were found to be genetically correlated with ASD [20]. Nordsletten et al. [21] found evidence of non-random mating within and across 11 psychiatric disorders including ASD in a Swedish population that examined correlation in diagnostic status. Additionally, studies have also investigated the presence of ASD-like traits in parents of ASD offspring [22–25]. It has been noted by many studies that parents of ASD offspring often present with ASD-like characteristics more so than expected (especially in multiplex (more than one affected individual per family) ASD families) [24,26-29].

There is evidence that assortative mating influences genotype frequencies that are associated with complex traits [24]. If assortative mating is present in a trait, and is not accounted for in a genetic study, it can confound heritability estimates [14,24,30]. Evidence from Klei et al. [31] also supports this as they found an elevated level of heritability in pseudo-controls (non-transmitted alleles which would be expected to behave as unaffected offspring) from multiplex ASD families. If parents are more genetically similar at casual variants for ASD (i.e. positive assortative mating is taking place) then the untransmitted alleles from the parents at these loci are more likely to be risk alleles and increase heritability estimates in pseudo-controls. Although it has been argued that assortative mating may only lead to a modest

increase of heritability estimates for a disorder [14].

Another concern is the assumption of random mating that is used when investigating genetic associations in ASD, which may be violated if assortative mating is taking place. Understanding the genetic aetiology of ASD may be completely entwined with the understanding of assortative mating [32].

To examine genetic assortative mating in ASD, we investigated the genetic similarity of the parents (spouses) compared to the genetic similarity of non-spouse pairings (we created) restricted to male/female pairings from the same ancestral population using kinship coefficients. The investigation of the genetic similarity between ASD parents was carried out using Single Nucleotide Polymorphism (SNP) data from Genome-Wide Association Studies (GWASs This approach has previously been used to examine assortative mating for traits such as height, BMI, education [5,7], but not to our knowledge in the context of ASD. An approach that has some similarity was used to investigate genetic assortative mating in ASD using pairwise genetic distance, in contrast to kinship coefficients, to investigate if spouses were more related than randomly mated parents [33]. Population stratification was not accounted for when estimating pairwise genetic distance but comparisons were within a small subset of the data obtained from one population. Another study also examined ASD genetic assortative mating using summary statistics from an ASD GWAS study but found no significant evidence due to sample size and limitations of ASD known variants [8].

We also investigated ancestral assortative mating using Principal Components Analysis (PCA) and phenotypic assortative mating using ASD trait data of the parents. We carried out our analyses on two ASD GWAS datasets, the Autism Genome Project (AGP) dataset [34,35] and the Simons Simplex Collection (SSC)

dataset [36].

Methods and Materials

Data

The AGP GWAS trio family dataset was collected at sites across Europe and North America and is described elsewhere [34,35]. Here we considered the Stage 2 dataset consisting of 2,931 multiplex (multiple affected individuals per family) and simplex (only one affected individual per family) families.

The SSC GWAS consists of data on 2,591 simplex North American families See [36-38] for further details on the SSC data. The majority of AGP and the SSC samples are of European ethnicity, see Figures S17 and S18 in Supplementary Information (SI) for PCA plots of the offspring.

The AGP GWAS data includes families grouped into two nested diagnostic categories, Strict ASD (autism diagnoses on both ADI and ADOS instruments [39,40]) and Spectrum ASD (autism-spectrum diagnoses on either the ADI or ADOS instruments), as defined in [35]. We applied the same ASD phenotype criteria as was used in the AGP [35] to define a Strict ASD phenotype within the SSC data. The Strict phenotype would be expected to have less clinical heterogeneity and therefore, have the potential to increase the power of identifying robust findings when compared to a broader autism diagnosis [41].

We also examined genetic assortative mating in a smaller sample comprising of the Autism Genetic Resource Exchange (AGRE) [42]. Due to the size of the dataset that was available for analysis and that we were unable to filter families on ASD diagnostic criteria details of the analysis of this dataset are available in the SI.

Quality Control Procedures

The Quality Control (QC) procedures follow a standard approach to trio GWAS QC, with individuals and SNPs removed when missingness > 0.05, Hardy Weinberg Equilibrium (HWE) p-value < 0.00001, and Minor Allele Frequency (MAF) < 0.05, see Table S1 in SI. We removed families that were related, and individuals with extreme levels of heterozygosity (greater than 2 standard deviations). We also excluded certain Linkage Disequilibrium (LD) ranges that can result in confounding in certain analyses such as PCA when examining population stratification [43]. We limited our analyses to complete trios as we needed both parents of the affected offspring for our analyses. After QC, the AGP with a Strict phenotype contained 1,590 trios and 712,319 SNPs and the SSC with a Strict phenotype contained 1,962 trios and 417,809 SNPs.

Statistical Methods

We used kinship coefficients to examine assortative mating in ASD. A kinship coefficient is the probability that two alleles sampled at random from two individuals are Identical-By-Descent (IBD occurs when two DNA sequences are inherited from a common ancestor). We compared the distribution of the kinship coefficients for spouses (mother/father pairings) with the distribution of all other possible non-spouse pairings restricted to male/female pairings from the same ancestral background, which we will refer to as all Non-Spouse Pairs. This is similar to the approach in [5] although we take into account population stratification using a different method. The Non-Spouse Pairs were restricted to male/female pairings as this is an investigation of genetic assortative mating between parents of offspring with ASD and genetic assortative mating may be different for same sex couples [5]. We filtered the Non-Spouse Pairs on ancestral population as creating non-spouse pairings across different ethnicities will result in individuals that look less genetically similar when comparing

them to spouses from the same ethnicity. This will lead to false positive findings for genetic assortative mating, see [44] and Section S1 in the SI.

The first step in our approach was to use the ADMIXTURE software [45] to estimate the amount of admixture (when individuals from genetically different populations mate and produce offspring) in the samples with the 11 populations in the HapMap3 dataset [46] as a reference (–supervised). Note the HapMap3 dataset used has had individuals removed due to cryptic relatedness (see [47] for further details). This allows us to identify the different populations that are contained within the ASD datasets. We then removed spouses that did not mate within the same ancestral population and only spouse pairs that had similar proportions of ancestry to each other and to others in their population remained (within 2 standard deviations of the mean of the proportion for each ancestral population).

In the second step, within these ancestral populations, we compared the distribution of the spouses' kinship coefficients, to the distribution of the Non-Spouse Pairs' kinship coefficients, where all kinship coefficients were estimated using the King software [48] (see Section S1 in SI). In this software, the kinship calculation accounts for heterogeneity between samples (for both the spouses and Non-Spouse Pairs see S1 in the SI). The quantiles (from 0.001 to 0.999 in increments of 0.001) for the spouse pairs' kinship coefficients calculated and then mapped to the kinship coefficients for the Non-Spouse Pairs, results were plotted in the same manner as Domingue et al. [5]. The 45° line indicates the null hypothesis that the genetic similarity among spouse pairs matches the genetic similarity among Non-Spouse Pairs. If the genetic similarity among spouses differs from the genetic similarity of Non-Spouse Pairs, then this is captured by departure from the 45° line and we calculate the area (and 95% Cis using 1,500 bootstrap replications) between this curve

and the 45° line. .

Before analysing the ASD datasets, we tested this approach in a subset of the HapMap dataset [49] which contains 101 spouse pairs from 3 different populations, HapMap Spouse dataset (see Section S1 in the SI). For this dataset, when fully accounting for the population stratification, we found no significant evidence of genetic similarity among the spouses as would be expected, see Section S1 in the SI.

The next approach investigated ancestral assortative mating using the method of Sebro et al. [3]: examining the correlation between the Principal Components (PCs) from the PCA of the genetic datasets. The PCs reflect the population structure within a dataset due to ancestry [50] and the estimation of correlation between the spouses' PCs reflects the degree to which the spouses are mating within their ancestral populations [3]. PCAs were carried out using Eigenstrat [51] on pruned sets of SNPs (using PLINK software with a window size of 50 SNPs, with the window shift set to 5 and R^2 threshold of 0.25) with high call rates greater than 0.999 (AGP dataset 124,547 SNPs and SSC dataset 115,734 SNPs after pruning).

We also investigated assortative mating among the parents using ASD trait information obtained from the Broad Autism Phenotype Questionnaire (BAPQ) (specifically subscales: *Aloof, Rigid, Pragmatic Language* and *Total BAPQ*) [52] and Social Responsive Scale - Adult version (SRS-A) (specifically subscales: *Awareness, Cognition, Mannerisms, Motivation*, and *Communication,* and the combined *Total SRS* score) [53]. The BAPQ includes both self and informant report versions which can be combined to give a more accurate result (*Best Score*), whereas the SRS-A is generally an informant report for the spouse or partner. For the SSC dataset, BAPQ data is available for 1,946 (99.2%) and SRS-A data is available for 1,958 (99.8%) of the spouses. Unfortunately, for the AGP data only 275 (17.3%) spouses have

complete BAPQ data and 428 (26.9%) spouses have SRS-A data.

Results

The Autism Genome Project

We investigated the genetic similarity of the parents in the AGP data by comparing the kinship coefficients of the spouses with the kinship coefficients of all other possible Non-Spouse Pairs. We required that both individuals in each spouse pair and Non-Spouse Pair were from the same ancestral population to avoid spurious results of assortative mating. We removed spouses that did not mate within the same ancestral population, resulting in 1,092 spouse pairs that belonged to one of six populations (see Figure S10 in SI).

The genetic similarity between spouse pairs compared to Non-Spouse Pairs is shown in Figure 1. Here the intersection of the vertical and horizontal lines represents where the median value (vertical line) of genetic similarity among spouses corresponds to the 0.52996 quantile (horizontal line) of all other possible Non-Spouse Pairs (within ancestral populations). The interpretation of this is that spouses are more genetically similar than all possible Non-Spouse Pairs. To calculate the degree of this increase in similarity among spouses, we calculated the area of the shaded region above the 45° line (0.0251, 95% CI = (0.0114, 0.0384). As this 95% CI does not contain 0 (the null hypothesis is that the area is 0, i.e. no assortative mating), this shows significant evidence that positive genetic assortative mating could be taking place in the AGP dataset.

INSERT FIGURE 1 ABOUT HERE

Collection site information was available for the AGP spouses, and we undertook an additional analysis where we compared spouses to Non-Spouse Pairs within the same site and ancestral population, see Figure 2. Here the genetic relatedness estimate

at the 0.5 quantile (vertical line) of spousal pairs corresponds to the 0.5576 quantile (horizontal line) of all Non-Spouse Pairs within site and ancestral population. The shaded area gives an estimate of assortative mating and is equal to 0.0451 here (95% CI = (0.0310, 0.0603)). This also indicates that the spouses are more genetically similar than would be expected.

INSERT FIGURE 2 ABOUT HERE

We also examined ancestral assortative mating in the AGP dataset by estimating the correlation between the spouses' PCs [3]. Figure 3 shows strong correlations between mothers and fathers for PC1 (this PC separates the Europeans from the Non-Europeans, see Figure S17 in SI for a PCA plot of the offspring) at 0.843 (p-value < 0.0001). This shows significant evidence for spouse pairs mating within their ancestral populations and hence, showing evidence of ancestral assortative mating.

INSERT FIGURE 3 ABOUT HERE

When investigating the correlation between mothers and fathers for ASD traits using the BAPQ and SRS-A, we found significant correlations between the parents on all ASD subscales for the AGP data apart from the *Aloof* subscale from the BAPQ and the *Motivation* subscale from the SRS (see Table ST2 and ST3 in the SI).

The Simons Simplex Collection

The same procedures carried out on the AGP dataset were carried out on the SSC dataset. As we wanted to compare spouses to all possible Non-Spouse Pairs in the SSC dataset, we again removed spouses that did not mate within the same population, resulting in 1,221 spouse pairs that belonged to one of six populations (see Figure S11 in SI). We used the kinship coefficients to estimate the genetic similarity between spouse pairs compared to all Non-Spouse Pairs, see Figure 4. The genetic relatedness estimate at the 0.5 quantile (vertical line) of spousal pairs corresponds to the 0.5032

quantile (horizontal line) of all Non-Spouse Pairs. The shaded area gives an estimate of assortative mating and is equal to -0.0062 (95% CI = (-0.0180, 0.0061)). Unlike for the AGP data, we see a negative value for the area indicating that the spouses are less similar than the Non-Spouse Pairs but this is not significant, therefore there is no evidence of assortative mating in the SSC dataset using this method. There was no site data available for the SSC to conduct any site related analyses.

INSERT FIGURE 4 ABOUT HERE

We examined ancestral assortative mating in the SSC by analysing the correlation between PC1 for the mothers and PC1 for the fathers (again, as for the AGP dataset, this PC separates the Europeans from the Non-Europeans, see Figure S18 in SI for PCA of the offspring). Figure 5 displays a strong correlation of 0.802 (p-value < 0.0001) between the mothers and fathers. This shows significant evidence of ancestral assortative mating, similar to the results for the AGP dataset.

INSERT FIGURE 5 ABOUT HERE

When investigating the correlation between mothers and fathers for ASD traits using the BAPQ in the SSC dataset, there was no evidence of a correlation, see Figure S24 in SI. With the SRS-A, we found significant correlations between the parents on all subscales bar the *Awareness* and *Motivation* subscales for the SSC data although the correlation values are lower than the AGP's SRS-A results. See Table ST4 in the SI.

Discussion

Assortative mating has been hypothesised to play a role in ASD, suggesting that people with ASD-like traits mate with one another more frequently than would be

expected. There has been evidence to suggest assortative mating in ASD can affect the prevalence and the heritability estimates [14, 31].

We also tested this kinship coefficient approach on the HapMap Spouse dataset, to assess whether we could account for population structure and avoid finding false evidence of assortative mating. We found no significant results for assortative mating in the HapMap Spouse dataset when restricting the non-spouse pairings to male/female pairings within their ancestral population (see Section S1 in the SI). If we did not take into account the population substructure correctly in this dataset, this led to very strong findings of assortative mating in the HapMap Spouse dataset (see Section S1 in the SI). [44]. We do note that this dataset is relatively smaller than the two ASD datasets analysed here and hence would be expected to have lower power to detect such effects. In addition, the ancestral populations are known in the HapMap Spouse dataset unlike in the AGP and SSC datasets.

We examined kinship coefficients to investigate genetic assortative mating in the AGP and the SSC datasets. We found significant evidence that the kinship coefficients for spouses were more similar when compared to those of Non-Spouse Pairs within the same ancestral population in the AGP data. Although we did not find any evidence of this in the SSC data.

Due to the spouse pairs and the Non-Spouse Pairs needing to be from the same ancestral population for our analyses, retaining only spouse pairs that had similar proportions of ancestry to each other and to others in their population, reducing our sample size. In particular, in the SSC dataset, 741 spouse pairs were removed (compared to 498 in the AGP). This, we suspect, will have reduced the power to detect the genetic similarity among the spouse pairs in the SSC dataset. Having stated this, when assortative mating is detected, as in the AGP dataset, it is less likely to be

confounded with population stratification, although we cannot fully rule this out.

Other populations have shown evidence of ancestral assortative mating when spousal correlation of the PCs on the genetic data has been investigated [3,28,54]. We also present evidence of this in the two ASD datasets by identifying a significant correlation between the spouses' PCs.

When investigating the ancestral assortative mating, it is impossible to try and tease apart how much of the spousal correlation is attributed to proximity of the spouses with each other. For instance in the AGP dataset, which is collected at many sites across Europe and North America, some countries will have less admixture than others and it is more likely that individuals from these populations will mate with other individuals from the same ancestral background based on proximity. The SSC data, on the other hand, was only collected at sites in North America, which has more admixture between different ethnicities, yet we still see strong spousal correlations on the PCs.

For the AGP dataset (where collection site data was available, not available for the SSC data), we randomly paired the spouses within ancestral population and collection site, as a proxy for proximity, obtaining significant findings here also (see Figure 3). This gives us more confidence that we are accounting for as much of the population stratification as possible, but we acknowledge that subtle population stratification may still be present.

We found no significant evidence of assortative mating in the SSC dataset using the kinship coefficients approach. In addition, for the phenotypic analyses, examining the ASD traits in the parents through the BAPQ and SRS-A, the findings in the SSC were not as strong as for the AGP dataset. These findings could be due to the differences in the ascertainment for the AGP and the SSC datasets. It is worth noting

that the SSC dataset had strict recruitment criteria, only including simplex families (only one individual with ASD per family) in the study. Furthermore, the parents of these families were additionally screened in the recruitment process for ASD traits, a design which inherently enriches for rare and de novo mutations [36]. The AGP had no such criteria and contains both multiplex and simplex families (approximately 38% of families are simplex [31]). Previous studies have investigated the correlation between ASD parents for ASD traits using these instruments and our results for the AGP data show similar correlations of 0.4 [25]. This is in contrast to results obtained for the SSC, where the SRS-A correlation values were much lower and there was no sign of a linear relationship between the parents for the BAPQ data (see Figure S24 in the SI).

Evidence has been shown that differences in the heritability estimates for ASD between multiplex and simplex families exist [31]. The results from Klei et al. and the stronger evidence of assortative mating in the AGP datasets that we have shown here, indicate that the genetic mechanisms differ between multiplex families and simplex families [31,55,56]. Understanding these differences, and the effects of the broader autism phenotypes present in parents of ASD offspring, will be imperative for understanding the etiology of ASD [28].

We acknowledge that the methods used here are not the only possible approaches to investigate assortative mating in a population. Other methods such as Polygene Risk Scores (PRS) could offer a means of exploring assortative mating among the parents for ASD risk variants. However, due to currently still relatively small sample sizes being available for ASD genotype studies, evidence for genetic variants associated with ASD is limited. For instance, the largest GWAS to date in ASD identified few additional findings [57]. This, coupled with the nature of this complex

disorder having many variants with small effect sizes [35], suggests that an approach using PRS would currently lack the power needed to find significant evidence. A previous study encountered these issues when trying to use such an approach in ASD [8]. However, such an approach may, in the future, offer another avenue for exploring assortative mating in ASD.

In conclusion, we found evidence to suggest genetic assortative mating is taking place in the AGP datasets and that there is no evidence of this when investigating a simplex family cohort, the SSC dataset. We have also identified significant evidence of correlations between parents with certain ASD traits in both the AGP and SSC datasets. Although this evidence of phenotype assortative mating is weaker in the SSC dataset. Further investigations are warranted into assortative mating in ASD as it can confound heritability estimates and increase prevalence estimates of the disorder. This study also further emphasises the different etiology that may be taking place between simplex and multiplex ASD families. It would certainly be of interest to investigate assortative mating at SNPs associated with ASD traits, although due to the complexity of ASD, the literature is not yet available to support this work, but we anticipate that this will be possible in the near future.

Acknowledgements

The authors are grateful to all of the families for participating in the AGP, as well as the principal investigators, AGRE and the AGP funders: Autism Speaks (USA), the Medical Research Council (UK), the Health Research Board (Ireland), the National Institutes of Health (USA), Genome Canada, the Canadian Institutes of Health Research and the Hilibrand Foundation. The AGP data sets can be obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap. The authors are grateful to all of the families at the participating SSC sites, as well as the principal investigators (A

Beaudet, R Bernier, J Constantino, E Cook, E Fombonne, D Geschwind, R Goin-Kochel, E Hanson, D Grice, A Klin, D Ledbetter, C Lord, C Martin, D Martin, R Maxim, J Miles, O Ousley, K Pelphrey, B Peterson, J Piggot, C Saulnier, M State, W Stone, J Sutcliffe, C Walsh, Z Warren and E Wijsman), and appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population data set described in this study by applying at https://base.sfari.org. In particular, we acknowledge S Sanders for his help with the SSC data. We acknowledge support from the Autism Genetic Resource Exchange (AGRE) and Autism Speaks. We gratefully acknowledge the resources provided by AGRE and the participating AGRE families. The Autism Genetic Resource Exchange (AGRE) is a program of Autism Speaks. "We acknowledge the support of the Trinity Centre for High Performance Computing. SC acknowledges the funding received from a Trinity College Dublin Studentship Award. We are grateful to reviewers for very helpful comments and suggestions.

Conflict of Interest

The authors have no potential conflicts of interest.

References

[1] Schneider KA and Peischl S (2011). Evolution of assortative mating in a population expressing dominance. PloS one, 6: e16821.

[2] De S, Pal SK, and Ghosh A (1998). Genotypic and phenotypic assortative mating in genetic algorithm. Information Sciences, 105: 209–226.

[3] Sebro R, Hoffman TJ, Lange C, Rogus JJ, and Risch NJ (2010). Testing for non-random mating: evidence for ancestry-related assortative mating in the Framingham heart study. Genetic epidemiology, 34: 674–9.

[4] Génin E, Ober C, Weitkamp L, and Thomson G (2000). A robust test for assortative mating. European journal of human genetics : EJHG, 8: 119–24.

[5] Domingue BW, Fletcher J, Conley D, and Boardman JD (2014). Genetic and educational assortative mating among US adults. Proceedings of the National Academy of Sciences of the United States of America, 111: 7996–8000.

[6] Conley D, Laidley T, Belsky DW, Fletcher JM, Boardman JD, and Domingue BW (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. Proceedings of the National Academy of Sciences, 113: 6647–6652.

[7] Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, et al. (2017). Genetic evidence of assortative mating in humans. Nature Human Behaviour, 1:0016.

[8] Yengo L, Robinson MR, Keller MC, Kemper KE, Yang Y, Trzaskowski M, et al. (2018). Imprint of Assortative Mating on the Human Genome. Nature Human Behaviour 2:948–954

[9] Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. (2014). Most genetic risk for autism resides with common variation. Nature genetics, 46: 881–885.

[10] Baron-Cohen S (2006). Two new theories of autism: hyper-systemising and assortative mating. Archives of disease in childhood, 91: 2–5.

[11] Williams JG, Higgins JPT, and Brayne CEG (2006). Systematic review of prevalence studies of autism spectrum disorders. Archives of disease in childhood, 91: 8–15.

[12] Harrison MJ, O'Hare AE, Campbell H, Adamson A, and McNeillage J (2006). Prevalence of autistic spectrum disorders in Lothian, Scotland: an estimate using the "capture-recapture" technique. Archives of disease in childhood, 91: 16–9.

[13] Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators (2012). Prevalence of autism spectrum disorders – Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. Morbidity and mortality weekly report Surveillance summaries (Washington, DC : 2002), 61: 1–19.

[14] Peyrot, W. J., Robinson, M. R., Penninx, B. W. J. H., & Wray, N. R. (2016). Exploring boundaries for the genetic consequences of Assortative mating for psychiatric traits. JAMA Psychiatry, 73(11), 1189–1195. http://doi.org/10.1001/jamapsychiatry.2016.2566

[15] Baron-Cohen S and Hammer J (1997). Parents of Children with Asperger Syndrome: What is the Cognitive Phenotype? Journal of cognitive neuroscience, 9: 548–54.

[16] Silberman S (2001). The geek syndrome. Wired Magazine.

[17] Baron-Cohen S, Wheelwright S, Burtenshaw A, and Hobson E (2007). Mathematical Talent is Linked to Autism. Human nature (Hawthorne, NY), 18: 125–31.

[18] Roelfsema MT, Hoekstra RA, Allison C, Wheelwright S, Brayne C, Matthews FE, and Baron-Cohen S (2012). Are autism spectrum conditions more prevalent in an informationtechnology region? A school-based study of three regions in the

Netherlands. Journal of autism and developmental disorders, 42: 734–9.

[19] Dickerson AS, Pearson DA, Loveland KA, Rahbar MH, and Filipek PA (2014). Role of parental occupation in autism spectrum disorder diagnosis and severity. Research in autism spectrum disorders, 8: 997–1007.

[20] Warrier, V., Toro, R., Chakrabarti, B., Group, the iPSYCH-B. autism, Børglum, A. D., Grove, J., … Kingdom. (2017). Systemizing is genetically correlated with autism and is genetically distinct from social autistic traits, 44(0), 1–22.

[21] Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, and Mataix-Cols D (2016). Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. JAMA psychiatry, 73: 354–61.

[22] Baron-Cohen S (2006). The hyper-systemizing, assortative mating theory of autism. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 30: 865–872.

[23] Happé F, Briskman J, and Frith U (2001). Exploring the cognitive phenotype of autism: weak "central coherence" in parents and siblings of children with autism: II. Real-life skills and preferences. Journal of child psychology and psychiatry, and allied disciplines, 42: 309–316.

[24] Constantino JN and Todd RD (2005). Intergenerational transmission of subthreshold autistic traits in the general population. Biological Psychiatry, 57: 655–660.

[25] Lyall K, Constantino, JN, Weisskopf MG, Roberts AL, Ascherio A, and Santangelo SL (2014). Parental social responsiveness and risk of autism spectrum disorder in offspring. JAMA Psychiatry, 71(8), 936–42. http://doi.org/10.1001/jamapsychiatry.2014.476

[26] Gerdts JA, Bernier R, Dawson G, and Estes A (2013). The broader autism phenotype in simplex and multiplex families. Journal of autism and developmental disorders, 43: 1597–605.

[27] Virkud YV, Todd RD, Abbacchi AM, Zhang Y, and Constantino JN (2009). Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics, 150B: 328–34.

[28] Page J, Constantino JN, Zambrana K, Martin E, Tunc I, Zhang Y, et al. (2016). Quantitative autistic trait measurements index background genetic risk for ASD in Hispanic families. Molecular Autism, 7(1), 39.

[29] Rubenstein E and Chawla D (2018). Broader Autism Phenotype in Parents of Children with Autism: A Systematic Review of Percentage Estimates. Journal of Child and Family Studies, 27(6).

[30] Sebro R, Peloso GM, Dupuis J, Risch NJ, Freedman M, Groop L, et al. (2017). Structured mating: Patterns and implications. PLOS Genetics, 13: e1006655.

[31] Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, et al. (2012). Common genetic variants, acting additively, are a major source of risk for autism. Molecular autism, 3: 9.

[32] Brosnan M and Walker I (2009). A preliminary investigation into the potential role of waist hip ratio (WHR) preference within the assortative mating hypothesis of autistic spectrum disorders. Journal of Autism and Developmental Disorders, 39: 164–171.

[33] Zhu, Z., Lu, X., Yuan, D., & Huang, S. (2017). Close genetic relationships between a spousal pair with autism-affected children and high minor allele content in cases in autism-associated SNPs. Genomics, 109(1), 9–15.

[34] Anney R, Klei L, Pinto D, et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet, 19: 4072–4082.

[35] Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Human molecular genetics, 21: 4781–92.

[36] Fischbach GD and Lord C (2010). The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron, 68: 192–5.

[37] Sanders SJ, Ercan-Sencicek GA, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. (2011). Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. Neuron, 70: 863–885.

[38] Chaste P, Klei L, Sanders SJ, Hus V, Murtha MT, Lowe JK, et al. (2014). A genomewide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? Biological Psychiatry.

[39] Lord C, Rutter M, and Le Couteur A (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of autism and developmental disorders, 24: 659–85.

[40] Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of autism and developmental disorders, 30: 205–23.

Biological Psychiatry.

[41] Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P, et al. (2001). The Autism Genetic Resource Exchange: A Resource for the Study of Autism and Related Neuropsychiatric Conditions. The American Journal of Human Genetics, 69(2), 463–466.

[42] Waltes R, Gfesser J, Haslinger D, Schneider-Momm K, Biscaldi M, Voran A, et

al. (2014). Common EIF4E variants modulate risk for autism spectrum disorders in the high-functioning range. Journal of neural transmission (Vienna, Austria : 1996), 121:1107–16.

[43] Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, et al. (2008). Long-range LD can confound genome scans in admixed populations. American journal of human genetics, 83: 132–5; author reply 135–9.

[44] Abdellaoui A, Verweij KJH, and Zietsch BP (2014). No evidence for genetic assortative mating beyond that due to population stratification. Proceedings of the National Academy of Sciences of the United States of America, 111: E4137.

[45] Alexander DH, Novembre J, and Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19: 1655–64.

[46] Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature, 467: 52–8.

[47] Stevens EL, Baugher JD, Shirley MD, Frelin LP, and Pevsner J (2012). Unexpected relationships and inbreeding in HapMap phase III populations. PloS one, 7: e49575.

[48] Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, and Chen WM (2010). Robust relationship inference in genome-wide association studies. Bioinformatics (Oxford, England), 26: 2867–73.

[49] International HapMap Consortium (2003). The International HapMap Project. Nature, 426: 789–96.

[50] Lee S, Epstein MP, Duncan R, and Lin X (2012). Sparse principal component analysis for identifying ancestry-informative markers in genome-wide association studies. Genetic epidemiology, 36: 293–302.

[51] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, and Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics, 38: 904–9.

[52] Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. Journal of Autism and Developmental Disorders, 37(9), 1679–90. http://doi.org/10.1007/s10803-006-0299-3

[53] Constantino JN, Gruber CP: The Social Responsiveness Scale manual. Los Angeles, CA: Western Psychological Services; 2005.

[54] Abdellaoui A, Hottenga JJ, de Knijff P, Nivard MG, Xiao X, Scheet P, et al. (2013). Population structure, migration, and diversifying selection in the Netherlands. European journal of human genetics : EJHG, 21: 1277–85.

[55] Bernier R, Gerdts J, Munson J, Dawson G, and Estes A (2012). Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. Autism research : official journal of the International Society for Autism

Research, 5: 13–20.

[56] Chaste P and Leboyer M (2012). Autism risk factors: genes, environment, and gene-environment interactions. Dialogues in clinical neuroscience, 14: 281–92.

[57] Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, T. A. S. D. W. G. of T. P. G. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Molecular Autism, 8, 21.

Figure Legends

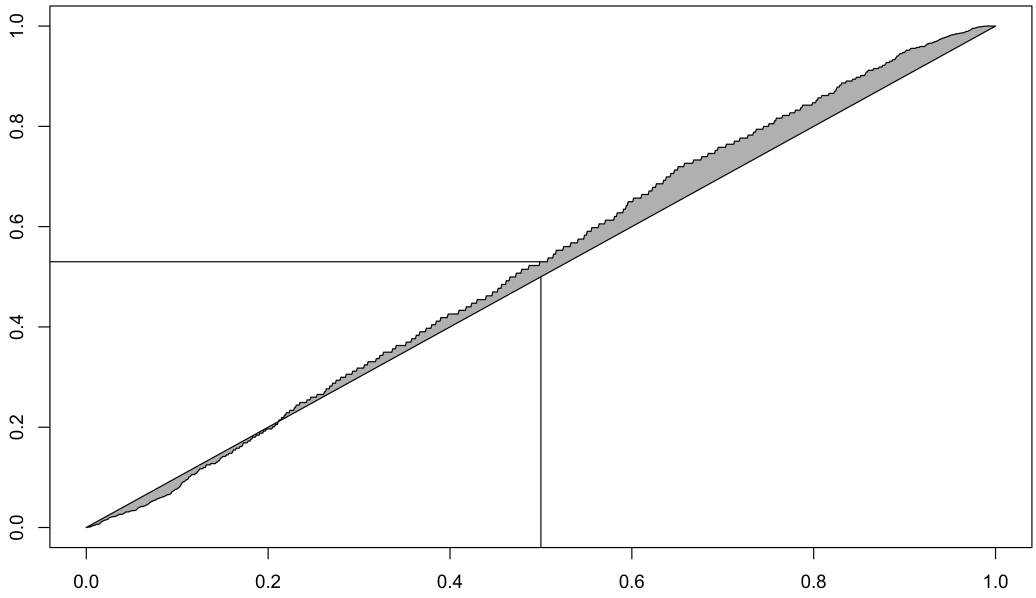
Figure 1: Assortative Mating in AGP dataset. The x-axis represents the quantiles of the distribution of kinship coefficients between spouse pairs. The y-axis represents quantiles of the distribution of kinship coefficients between all other Non-Spouse Pairs. The shaded area gives an estimate of assortative mating and is equal to 0.0251 here (95% CI = (0.0114, 0.0384)). The genetic relatedness estimate at the 0.5 quantile (vertical line) of spousal pairs corresponds to the 0.52996 quantile (horizontal line) of all other Non-Spouse Pairs.

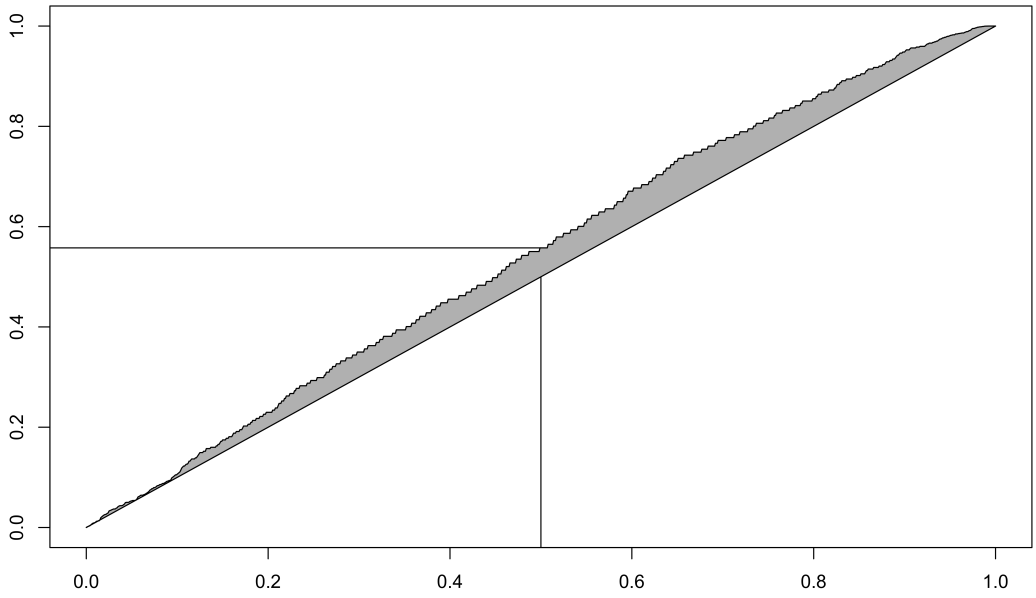
Figure 2: Assortative Mating in AGP dataset within site. The x-axis represents the quantiles of the distribution of kinship coefficients between spouse pairs. The yaxis represents quantiles of the distribution of kinship coefficients between all other Non-Spouse Pairs within the same site and ancestral population. The shaded area gives an estimate of assortative mating and is equal to 0.0451 here (95% CI = (0.0310, 0.0603)). The genetic relatedness estimate at the 0.5 quantile (vertical line) of spousal pairs corresponds to the 0.5576 quantile (horizontal line) of all other Non-Spouse Pairs.

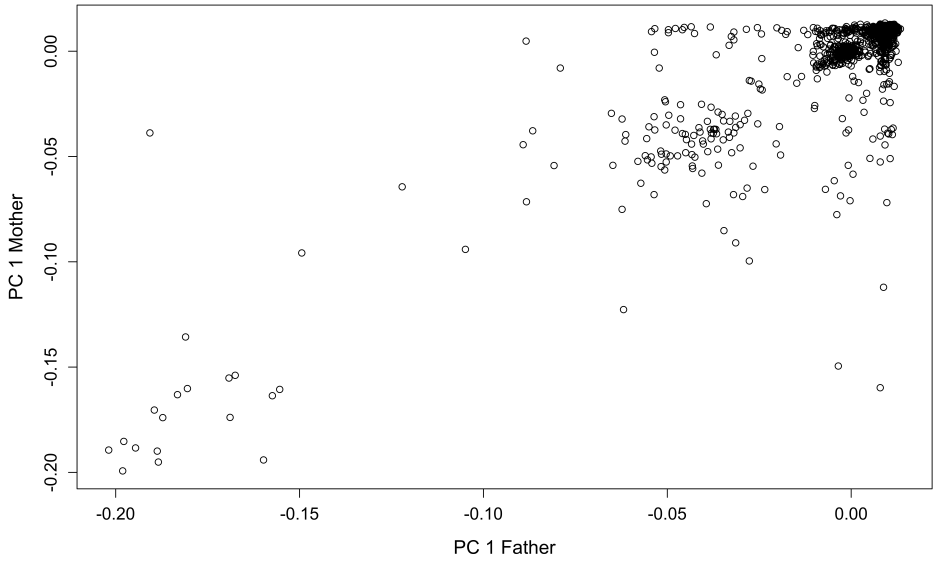
Figure 3: The PC 1 of Mothers versus Fathers in the AGP dataset. The spousal correlation for PC 1 is 0.843.

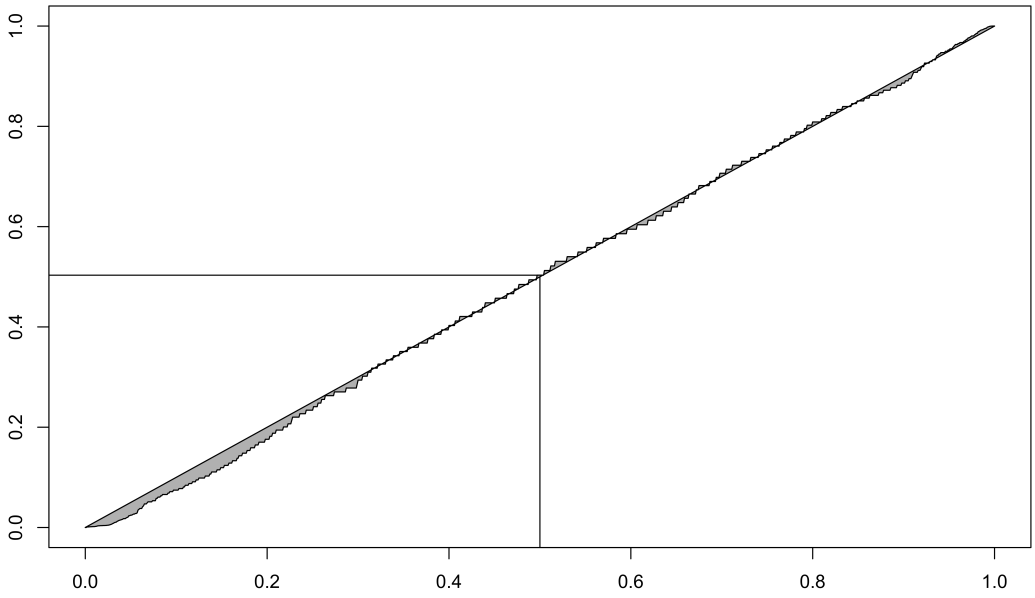
Figure 4: Assortative Mating in SSC dataset. The x-axis represents quantiles of the distribution of kinship coefficients between spouse pairs. The y-axis represents quantiles of the distribution of kinship coefficients between all other Non-Spouse Pairs. The shaded area gives an estimate of assortative mating and is equal to -0.00622 here (95% CI = (-0.0180, 0.0061)). The genetic relatedness estimate at the 0.5 quantile (vertical line) of spousal pairs corresponds to the 0.5032 quantile (horizontal line) of all Non-Spouse Pairs.

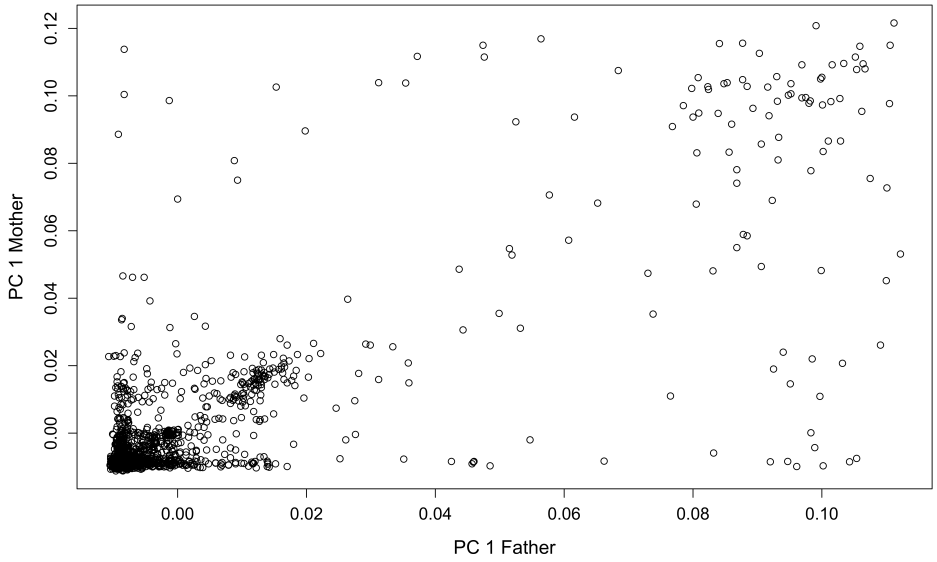
Figure 5: The PC 1 of Mothers versus Fathers in the SSC dataset. The spousal correlation for PC 1 is 0.802.


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


PC 1 Mother
0.00
-0.05
-0.10
-0.15
-0.20
-0.20
-0.15
-0.10
-0.05
0.00
PC 1 Father


0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


PC 1 Mother
0.00
0.02
0.04
0.06
0.08
0.10
0.12
PC 1 Father
0.00
0.02
0.04
0.06
0.08
0.10